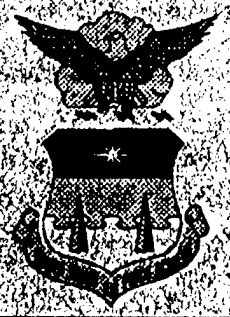


UNITED STATES AIR FORCE ACADEMY EDUCATIONAL OUTCOMES ASSESSMENT WORKING GROUP

PHASE 2:

FINAL REPORT

April 1997

REPORT DOCUMENTATION PAGE	Form Approved OMB No. 0704-0188

1. AGENCY USE ONLY *(Leave blank)*	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	30 April 1997	Outcomes Assessment

4. TITLE AND SUBTITLE
Educational Outcomes Assessment Working Group, Phase II, Final Report

5. FUNDING NUMBERS

6. AUTHOR(S)
Educational Outcomes Assessment Working Group,
Colonel David B. Porter, Chair (Ed).

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Educational Outcomes Assessment Working Group

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
Dean of the Faculty
HQ USAFA/DF
USAF Academy, CO 80840

10. SPONSORING/MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION AVAILABILITY STATEMENT
UL

12b. DISTRIBUTION CODE

13. ABSTRACT *(Maximum 200 words)*

This publication provides an account of educational outcomes assessment activity undertaken by seven assessment teams under the Phase II Charter of the Dean of the Faculty's Educational Outcomes Assessment Working Group. Data from thirty-five core academic courses were examined and rated as to their absolute contributions to each of three educational outcomes: integrated knowledge, critical thinking, and intellectual curiosity. Background instruments, raw data, team reports, assessed contribution ratings, and meta analyses are included as separate chapters. The final chapter includes conclusions and implications for unit self assessment, curricular reform, and further research.

14. SUBJECT TERMS	15. NUMBER OF PAGES
Educational Outcomes Assessment	201
	16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
Unclassified	U/C	U/C	UL

GENERAL INSTRUCTIONS FOR COMPLETING SF 298

The Report Documentation Page (RDP) is used in announcing and cataloging reports. It is important that this information be consistent with the rest of the report, particularly the cover and title page. Instructions for filling in each block of the form follow. It is important to ***stay within the lines*** to meet ***optical scanning requirements.***

Block 1. Agency Use Only *(Leave blank)*.

Block 2. Report Date. Full publication date including day, month, and year, if available (e.g. 1 Jan 88). Must cite at least the year.

Block 3. Type of Report and Dates Covered. State whether report is interim, final, etc. If applicable, enter inclusive report dates (e.g. 10 Jun 87 - 30 Jun 88).

Block 4. Title and Subtitle. A title is taken from the part of the report that provides the most meaningful and complete information. When a report is prepared in more than one volume, repeat the primary title, add volume number, and include subtitle for the specific volume. On classified documents enter the title classification in parentheses.

Block 5. Funding Numbers. To include contract and grant numbers; may include program element number(s), project number(s), task number(s), and work unit number(s). Use the following labels:

C - Contract	**PR** - Project
G - Grant	**TA** - Task
PE - Program Element	**WU** - Work Unit Accession No.

Block 6. Author(s). Name(s) of person(s) responsible for writing the report, performing the research, or credited with the content of the report. If editor or compiler, this should follow the name(s).

Block 7. Performing Organization Name(s) and Address(es). Self-explanatory.

Block 8. Performing Organization Report Number. Enter the unique alphanumeric report number(s) assigned by the organization performing the report.

Block 9. Sponsoring/Monitoring Agency Name(s) and Address(es). Self-explanatory.

Block 10. Sponsoring/Monitoring Agency Report Number. *(If known)*

Block 11. Supplementary Notes. Enter information not included elsewhere such as: Prepared in cooperation with....; Trans. of....; To be published in.... When a report is revised, include a statement whether the new report supersedes or supplements the older report.

Block 12a. Distribution/Availability Statement. Denotes public availability or limitations. Cite any availability to the public. Enter additional limitations or special markings in all capitals (e.g. NOFORN, REL, ITAR).

DOD -	See DoDD 5230.24, "Distribution Statements on Technical Documents."
DOE -	See authorities.
NASA -	See Handbook NHB 2200.2.
NTIS -	Leave blank.

Block 12b. Distribution Code.

DOD -	Leave blank.
DOE -	Enter DOE distribution categories from the Standard Distribution for Unclassified Scientific and Technical Reports.
NASA -	Leave blank.
NTIS -	Leave blank.

Block 13. Abstract. Include a brief *(Maximum 200 words)* factual summary of the most significant information contained in the report.

Block 14. Subject Terms. Keywords or phrases identifying major subjects in the report.

Block 15. Number of Pages. Enter the total number of pages.

Block 16. Price Code. Enter appropriate price code *(NTIS only)*.

Blocks 17. - 19. Security Classifications. Self- explanatory. Enter U.S. Security Classification in accordance with U.S. Security Regulations (i.e., UNCLASSIFIED). If form contains classified information, stamp classification on the top and bottom of the page.

Block 20. Limitation of Abstract. This block must be completed to assign a limitation to the abstract. Enter either UL (unlimited) or SAR (same as report). An entry in this block is necessary if the abstract is to be limited. If blank, the abstract is assumed to be unlimited.

Educational Outcomes Assessment Working Group Report
Phase II: Core Course Contributions
Table of Contents

Foreword

Over the last decade, the Air Force has moved to incorporate many of the principles of Total Quality Management. In 1991 the Air Force Academy initiated its quality transformation by acquainting all faculty members with quality principles and their application to education. Here at the Academy, one consequence of increased awareness of quality principles has been an increasing focus on institutional purpose; our development of seven specific educational outcomes reflects this emphasis. The quality paradigm has also affected academic processes at the Air Force Academy. We've increased our reliance on interdisciplinary forums and committees to provide feedback on all institutional processes and programs.

Our cadre from the initial total quality education effort became a standing committee in 1992 with separate subcommittees addressing internal and external stakeholders. The "internal" subcommittee later evolved and expanded to become the Educational Outcomes Assessment Working Group in the fall of 1994. Their discussions led the group to focus on three particular outcomes (written communication, the ability to frame and resolve ill-defined problems, and intellectual curiosity), to develop rubrics for each of these, to develop and test common assessment tasks and also to explore a variety of potential educational process metrics and instruments for their collection. This group sponsored and coordinated a variety of independent educational research projects. Phase I, Initial and Final reports extensively document the team's activities and accomplishments over these first 18 months.

The Phase II charter was very ambitious: it sought to assess all our core courses' contributions to all seven Air Force Academy Educational Outcomes. This final report is a first cut toward that end, and as such, a starting point rather than a definitive statement or a declaration of policy. Let me reiterate that the department head has the authority and responsibility to analyze and use the data they believe valid and significant in their curricular change processes. I am interested in assessing "critical thinking" and how you go about it in your departments. Consequently, departments are urged to incorporate the data in their Unit Self Assessments, and this report should be useful as a **guide**, not a mandate, in that process. I applaud the efforts of all the individuals who worked together to produce this report; it provides a unique and significant opportunity for us all to continue to improve our academic processes and programs.

RUBEN A. CUBERO
Brigadier General, USAF
Dean of the Faculty

Acknowledgments

Educational outcomes assessment is a team sport. This fact is perhaps best reflected in the names of the members of the various assessment teams that appear at the front of each of the reports. Although individuals varied in their prior assessment experience and expertise, all shared a willingness to put aside personal bias to strive to assess course contributions fairly. The chairs of each of the respective teams, in particular, undertook a great deal of responsibility for coordinating with members and seeing the reports through to completion. I'd also like to offer special recognition to Major Tony Aretz; although he did not serve on any one particular team, he selflessly provided his expertise and counsel to several of the separate teams on numerous occasions. He also was a willing reviewer of early drafts. Much of what is best about this report is likely to have come from one of Tony's ideas or casual comments.

I also received considerable feedback and editorial comments from many of the other team chairs and members. This report is much better for their inputs. Similarly, several other permanent professors took the time to provide substantive comments and criticisms of an original draft provided to the Faculty Council. In particular, Col Ron Reed of the Department of Biology and Col Jim Head of the Department of Physics provided comments which were insightful and provocative. They often persuaded me to change my mind as well as my words. However, I was not always persuaded and several controversial conjectures doubtless remain. Their intent is to provoke consideration of alternatives and conversations about what is best for our students; not just to antagonize those with other perspectives.

I also had the great good fortune of having the counsel and support of several individuals I most respect in the field of educational assessment. Even from the project's inception, these individuals provided the encouragement and expert advice required to undertake such a task. The enthusiastic reviews of an early draft provided by the first three individuals provided the validation I needed to see this project through to completion. These individuals are: Dr. Monica Manning, former Convener of the American Association of Higher Education's Annual Conference on Quality and Assessment and now Executive Officer of The Nova Group; Dr. Cecilia Lopez, Associate Director of the North Central Association of Colleges and Universities, and Dr. Ted Marchese, Vice President of the American Association of Higher Education and Editor of Change, the Magazine of Higher Learning. Additionally Peter and Noreen Facione, authors of the California Critical Thinking Dispositions Inventories and many other assessment instruments, ideas and approaches provided invaluable suggestions and the prototype for our measurement of classroom climate.

Finally, I'd like to recognize the three individuals who provided a thorough editorial review and expert technical assistance in collating and formatting the various data bases, team reports and comments. Capt Crystal Jonas' thorough proofing caught a multitude of errors. Mrs. Laura Neal and Capt Michele Fincher worked tirelessly to reformat and integrate materials as they became available. Clearly, "the look" and accessibility of the information contained in this document is due to their extraordinary efforts. Cheers,

DAVID B. PORTER, Col, USAF
Phase II, Assessment Team Chair

Chapter 1

Introduction and Background

CHAPTER ONE: INTRODUCTION & BACKGROUND

An Integrated Approach to Educational Outcomes Assessment

Published in the Journal of Adult Assessment, Volume VI, Number 2, (Summer, 1996).

David B. Porter
Sandra M. Eisenhut
United States Air Force Academy

Every system is perfectly designed to yield the results observed.
– Peter Sholtes

If you don't know where you're going, you'll probably end up somewhere else.
– David Campbell

Assessment refers to the activities undertaken by an organization to obtain the information it needs to improve its ability to produce the outcomes it values. Educational systems are designed to cause changes to occur within students (and less obviously: faculty). Ultimately, the effectiveness of any institution must be measured by the intellectual contributions to society made by those who are or have been a part of that institution. Even our largest institutions are imbedded within larger social systems which determine the true value of educational outcomes and provide the appropriate context for developing internal assessment activities. This chapter provides a brief historical account of the attempts by the United States Air Force Academy to integrate educational outcomes assessment with other processes inherent to an institution of higher education.

Institutional Context and Background

The United States Air Force Academy, located just north of Colorado Springs, Colorado is both an undergraduate, degree-granting, academic institution and also an operational Air Force organization. This double identity poses some unique challenges and creates some unique opportunities. The Academy's student population of approximately 4,000 "cadets" receive instruction from an academic faculty of just over 500 as well as an additional 100 military officers who administer a variety of military training programs. Cadets do not "pay" tuition, and in fact, actually "earn" half the base pay of an Air Force second lieutenant while at the Academy. Each year, cadets are competitively selected from among the top high school graduates across the nation. Their near 1250 average SAT scores, however, are not their most distinctive characteristic as a student body; over 80 percent earned letters in varsity sports in high school. Over eighty percent of the Academy's faculty are military officers, the majority of whom have earned Master's degrees only and are assigned to the Academy for a single three or four year military "tour". Pursuant to recent changes to public policy, the Academy is increasing the number of civilian faculty (nearly all of whom possess terminal degrees) toward a goal of 25 percent by the year 2000. As a military organization, the Academy faculty is very stable with an unusual degree of continuity provided by a few senior officers with relatively permanent status (three percent) and others allowed to extend their service beyond four years (an additional ten

percent) and its new civilian faculty positions. However, as an institution of higher education, the faculty's average college teaching experience of about two years suggests an educational system plagued by continual flux, predestined to perpetual intellectual immaturity and academic instability. As is often the case, the reality lies somewhere between these extreme views. Nonetheless, the Air Force Academy provided a unique context for the development of an integrated program of educational outcomes assessment.

The mission of the United States Air Force Academy is to *develop and inspire air and space leaders with vision for tomorrow*. The accomplishment of this mission challenges everyone to seek ways to change the things they do so they might make more frequent and substantial contributions to the development and inspiration of others assigned to the Academy. Knowing the extent to which this mission is being accomplished is important. Congress wants to know that taxpayers are getting their "money's worth", but even more importantly, each individual at the Academy is accountable to all other individuals to ensure that a system is created in which everyone has the opportunity and encouragement to develop and contribute. The Dean recognized that *to develop and inspire* students, faculty members must not only establish high academic and professional standards, they must also motivate students to meet these standards. In order for graduates to contribute meaningfully to the Air Force, they must *know* things they didn't know when they arrived at USAFA; they must be able to *do* things they couldn't do before and they must have positive *attitudes* toward themselves, their colleagues and the Air Force. Assessing the extent to which USAFA is achieving its mission is essential; failure to attempt to measure those things that are most important because they are too complex or ambiguous invites organizational distortion in the direction of readily available but functionally peripheral matters and metrics.

As a first step in this process, the Academy's most senior academic officers met throughout academic year 1992-1993 and eventually reached consensus on seven particular educational outcomes they agreed were critical indicators of the institution's educational success. In the summer of 1994, it became apparent that the best way to affirm the institution's commitment to these outcomes would be to begin the process of actually measuring them. This was the task assigned to the Educational Outcomes Assessment Working Group.

Phase I: Development of the Educational Outcomes Assessment Working Group

The Educational Outcomes Assessment Working Group was comprised of 20 faculty volunteers representing each of the four academic divisions (viz., Basic Sciences, Engineering, Humanities, and Social Sciences) as well as several support agencies. Most of the military members of the group were drawn from the ten percent of faculty granted continuing tours. The group focused on discovering ways to determine whether or not Academy graduates would be likely to make meaningful contributions in their service to the United States Air Force and the nation. More specifically, the group's initial objectives were to: 1) increase understanding of the ways in which students develop; 2) determine which educational activities made the greatest positive contributions to the development and inspiration of cadets; and 3) collect relevant in-process measures of educational activities to provide convergent support for the overall assessment of graduate attributes. Together, these objectives should suggest academic alternatives and help policy makers improve the educational system.

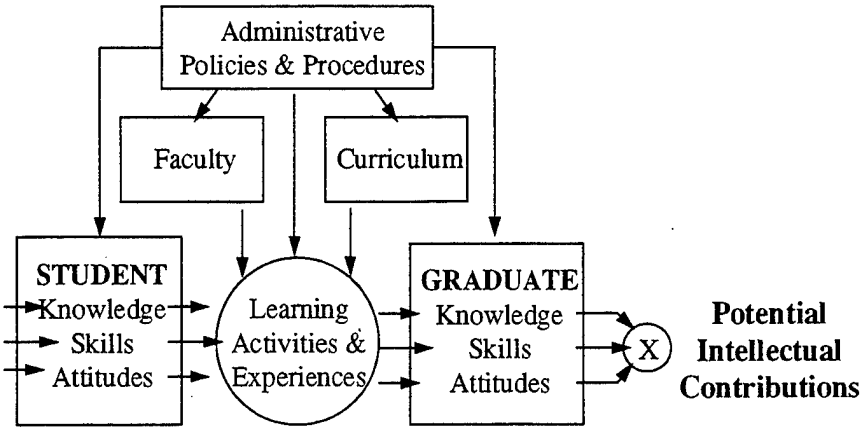


Figure 1. The General Education Model.

To address these needs, the group embarked on an ambitious self-education program. We struggled with the notion of what education was all about and how it worked. The group agreed on a general model which identified important components and suggested several causal relationships. This model is shown in Figure 1 above. The ultimate goal of any educational system is to produce graduates who will make substantive contributions to their community or society. These potential contributions are the product of what graduates know, what they can do and also their attitude toward themselves, their work and the community they serve. Changes occur along each of these developmental channels throughout the time a student is in the educational system. The notion that the subsequent potential is a *product* rather than a *sum* implies that extra emphasis or development in one particular area cannot compensate for deficiencies in another area. Graduates with vast knowledge and expert skill but negative attitudes will neither serve society nor reflect credit on their institution. From the perspective provided by this general model, the educational outcomes were seen as unique combinations of the knowledge, skills and attitudes desired in graduates. With the general model agreed upon (as well as general guidelines concerning assessment), the group focused on three particular outcomes: written communication, the ability to frame and resolve ill-defined problems, and the development of intellectual curiosity. The English Department had already developed evaluative criteria for cadet writing. The working group studied and revised these standards as a preliminary exercise for developing rubrics for the two other educational outcomes. The group gradually developed consensus concerning objective descriptions of the behaviors reflective of Excellent, Satisfactory or Deficient levels of student attainment of the ability to resolve ill-defined problems and an attitude of intellectual curiosity.

Ill-defined problems are those which necessarily do not have a single correct solution; they require assumptions to compensate for data deficits. Framing and resolving (i.e., re-solving) ill-defined problems is primarily a skill but involves knowledge and attitudinal components as well. The working group recognized that the assessment of this skill was itself an ill-defined problem. Piaget's work on early childhood development provided a useful theoretical foundation (i.e.,

frame) for exploring more contemporary theorists who've looked at late adolescent and adult development. In particular, Perry's work on intellectual and moral development of college students seemed to provide a useful extension and elaboration of Piaget's theory. The more recent work done by King and Kitchener in developing and testing their Reflective Judgment Model was even closer to our needs. This model suggests that thinking skills (viz., reflective judgment ability) develop in distinct stages. The criteria for an *Excellent* rating of students' framing and resolving ability converged with King and Kitchener's characterization of the approach taken by graduate students (i.e., contextual, incremental and insightful); *Satisfactory* criteria approximated transitional stages; and thinking that was pre-reflective (i.e., involving a dualistic reliance on authority and assumed absolutes) was deemed to be *Deficient.*

Having a rubric to provide criteria for evaluating examples of student problem solving was only the first step. Recognizing that ill-defined problems necessarily don't have single "right" answers, the group undertook the task of developing a suitably "ill-defined" problem with which to gather data on students' skill. Again Piaget provided a starting point. His classic "plant problem," used to distinguish formal from concrete operations in children was transformed to a relatively realistic aircraft deployment scenario. Students were initially given the ill-defined and incomplete problem to solve; then 12 minutes later, given a set of questions to answer about the process they used, their recommended solution and level of confidence, and what additional information might have improved the quality of their analysis. After several iterations of pilot studies, this task was administered to all 450 students enrolled in a mandatory senior level engineering course (general education requirements at USAFA include 7 engineering courses). Demographic data concerning GPA, gender and academic major were masked before committee members used the rubrics previously developed to evaluate responses on a scale ranging from 1(Deficient) to 5 (Excellent). A modified nominal group technique was used to train raters. Each rater evaluated several student responses and these ratings were then presented to the rest of the group; individuals openly discussed reasons for discrepancies. At the end of a couple one-hour sessions, individual ratings came to be amazingly consistent and reliable. All 450 responses were scored over the summer. Results suggested that most cadets displayed at least a satisfactory level of performance, but about 20 percent had produced deficient responses. There were no substantive differences attributable to gender, academic major or GPA. This last result was intriguing and at least a little disturbing. Did it mean that the assessment task was not valid or did it perhaps mean that GPA did not really reflect the ability to work with ill-defined problems? To address these questions we examined the data from students enrolled in an academic program specifically designed to address ill-defined problems (viz., Operations Research). For the 12 students in this major who completed the assessment task, the correlation between their major's GPA and task performance was .73. This finding supported the validity of the instrument but also legitimized concern about the extent to which most of our courses and programs involve the development of these important thinking skills.

The intellectual curiosity educational outcome posed another formidable assessment task. Whereas framing and resolving was mostly skill, intellectual curiosity was largely a matter of will (or attitude). However, with considerable debate, the group again was able to arrive at a general rubric broken into the three levels of attainment (*Excellent, Satisfactory or Deficient*). In a nutshell, *Excellent* intellectual curiosity was manifest by students whose interests were broad, deep and persistent; *Satisfactory* intellectual curiosity was shown by students whose work was

marked by some of these characteristics some of the time; and *Deficient* curiosity was reflected by the all too familiar student aphorism, "whatever" (i.e., the general absence of intellectual depth, breadth, persistence and responsibility). Although students' intellectual curiosity might be shown on a variety of class projects, an alternative approach was taken to its assessment. Two important questions were: What kinds of things do students observe or experience in the classroom that pique their curiosity? What practices and projects encourage students to wonder and conversely, what inhibits their curiosity? A senior student undertook these questions as the focus of an independent research project. After reviewing the educational psychology literature, he organized several student focus groups of freshmen and juniors; introduced the idea of intellectual curiosity to them, then recorded what these students had to say. Four general themes emerged: 1) for most students, personal *choice* is a powerful motivator (and absence of choice is an equally potent inhibitor); 2) students know when teachers don't understand the teaching and learning process; 3) there is a clear difference in activities which lead to learning and those that are used for evaluation and selection (i.e., tests); and 4) attitudes are important; the climate the teacher creates from the very first day can either encourage or quash curiosity.

The working group also looked at a variety of in-process measures. Based on the model presented earlier, teachers, curricula and administrative policies and procedures all affect the development of students' knowledge, skills and attitudes. One project involved developing an inventory of current curricular policies. Directors of thirty general education courses and a random sample of another forty courses completed both objective questions about their course policies and practices as well as providing a rich account concerning ways in which they believed their courses addressed the educational outcomes. Another project involved asking faculty directly about their teaching perspectives and practices along with a variety of demographic questions. A response rate of 58 percent permitted data to be aggregated by department and compared to average course critique ratings for each of the 19 academic departments. Somewhat surprisingly, more quizzes and exams were found to be the best predictor of high ratings of the overall course and the amount cadets thought they had learned. Many of us had suspected that regular feedback would enhance learning, but did not think that it would be reflected by student critiques. Another intriguing finding concerned the prediction of students' perception of teaching effectiveness. Ever since an institutional student course critique system was adopted, there has been concern that the predominantly male student body would not rate female faculty members as positively as they rated male faculty members (in fact, evidence of this bias has been found at many civilian universities as well). In this context, it was somewhat surprising to find that the *best* predictor of an academic department's average rated teacher effectiveness was the proportion of female faculty ($r^2 = .29$); departments with a larger proportion of female faculty were rated the most positively (academic departments ranged from 0 to 35% female faculty).

Another study sought to examine the relationship between individual faculty demographic characteristics, temperament, teaching philosophy and rated effectiveness. One hundred ten faculty members completed demographic questions, a Meyers BriggsTemperament Indicator, and a question about the relative emphasis they personally placed on each of the three educational development channels (i.e., students' knowledge, skills and attitudes). This study showed that many demographic variables had only negligible effects on effectiveness ratings (e.g., height, gender, aeronautical rating, terminal degree and USAFA alumni status). The best demographic predictor of positive ratings from students was the number of years teaching experience at the

Academy. The temperamental data suggested a significant advantage for intuitive instructors and a marked disadvantage for those whose temperament was more traditional or authoritarian. However, further analysis revealed that although temperament did an excellent job of predicting rated effectiveness differences between academic divisions (i.e., Humanities and Social Science faculty members were seen as being less traditional and Basic Sciences and Engineering faculties saw themselves as being more traditional), it was not a significant factor in predicting rated effectiveness within any of the four divisions. The effects of teaching perspective turned out to be about twice as great as either the demographic or temperamental predictor variables. Teachers who placed much more emphasis on students' knowledge than on their attitudes received low effectiveness ratings; those who placed approximately the same weight on attitudes and knowledge were consistently perceived as being more effective. This robust effect was observed both within and across academic divisions. Finally, these three factors (experience, temperament and perspective) were found to be independent (i.e., non interactive) influences which combined to explain 63% of the variance in teachers' rated effectiveness.

The activities of the Educational Outcomes Working Group epitomized continuous quality improvement and have substantially enhanced the understanding and effectiveness of educational processes at the Air Force Academy. Several aspects of this program were unique. First, the charter this group was given was very broad, complex and difficult given the nature of the Educational Outcomes themselves (outcomes were selected not because they would be easy to measure but because they were assumed to be essential and significant). Only one of the seven original educational outcomes involved specific knowledge content; the rest of the outcomes related to the general skills, abilities and attitudes. Second, the magnitude and scope of this effort extended across all academic departments at the institution. Traditional academic values such as skepticism and autonomy made the high degree of inclusion and integration particularly challenging. Third, success required the creation of a general and coherent framework for integrating multiple measures of outcomes and processes as well as diverse academic perspectives to provide a general assessment framework. A serious attempt was made to develop this program in a manner which was itself consistent with the principles of continuous quality improvement. Finally, the diversity of the group members was substantial; volunteers were drawn from different academic disciplines, different educational backgrounds, and different occupational experience (including both military and civilian). The overall structure of these assessment activities was designed to encourage improvement *without coercion or fear*. As a result, there was both a high level of faculty and student participation and a consistently high quality of contributions by group members. After two years of work, multiple metrics and methodologies were developed and refined; masses of data collected from students, faculty and administrators; multiple integrative analyses performed, reported and distributed; and the group members retained most of their original energy and enthusiasm.

The time invested by the Working Group in Phase I activities was estimated to be between 15 and 20 hours per month per committee member--a total of approximately 5,000 man-hours over the 18 months of Phase I assessment activities. In addition to planning time, assessment activities, and analyses, the Working Group conferred with numerous experts in education and assessment and also presented their assessment activities at the American Association of Higher Education Annual Conference and the North Central Association of Colleges and Universities' Annual Conference to share their findings with other educators. The Phase I Educational

Outcomes Assessment Working Group was composed of 20 individuals representing diverse academic constituencies. With the exception of two individuals whose promotions necessitated transfer to other geographical locations, all other members remained active.

Initial and Final Reports of Phase I activities each contained nearly 100 pages outlining what the Working Group had accomplished. Each member of the committee served as a point of contact with other faculty within her or his own academic division or agency. Members were committed to excellence in the assessment activities and personally invested in "selling" the program to their respective departments or agencies. As evidence for this support, 58 % of the faculty volunteered to participate in the Faculty Practices Survey--a very high level of participation. Nearly 20 % even volunteered to complete personal demographic, attitudinal and temperamental questionnaires and return them with private course and instructor critique information to permit more detailed analysis of the determinants of perceived teaching effectiveness. The uniqueness of the Academy-specific Educational Outcomes demanded that each component of the assessment program be customized: the faculty practices survey was written to fit the unique Academy situation; the student focus groups were designed to tap cadet perspectives and used protocols to minimize distortions created by differential military ranks, and the ill-defined problem was specifically tailored to provide a realistic Air Force context within which cadets could demonstrate their prowess. The group also addressed the general question of how educational assessment might best contribute to mission accomplishment and organizational renewal. A conceptual model showing the necessity of maintaining synchrony between the development of trust and understanding is shown in Figure 2. By their nature, assessment activities are likely to increase understanding; it is essential that as educational processes become more transparent, trust within the organization also be increased. This model served as an implicit framework for the groups' internal development as well as its interaction with the broader academic and military community at USAFA.

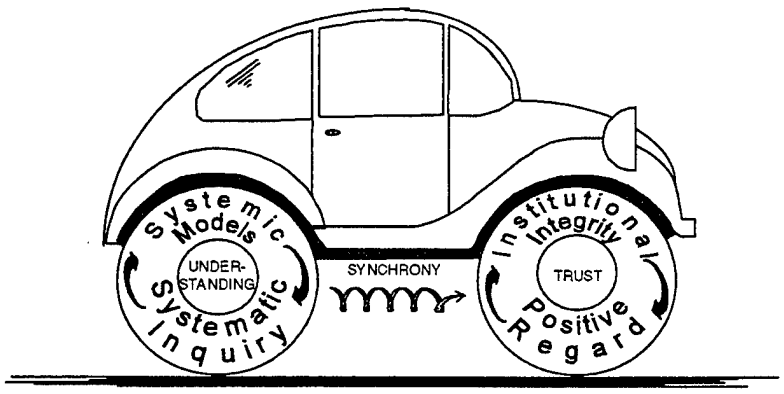


Figure 2. The Qualitymobile.

Beyond the development of assessment tools, a variety of auxiliary assessment methodologies were also developed by the Working Group. The importance of random and representative sampling procedures, standardized administration protocols, instrument analysis, improvement and validation and the criticality of student motivation and cooperation were all

discovered and rediscovered. Additionally, the value and potency of student participation in planning and conducting, as well as participating in educational research, were repeatedly demonstrated. From a broader context, the necessity to gradually integrate findings from diverse studies was recognized as being essential to sustaining organizational effectiveness. Like many other institutions, the Air Force Academy has collected a great deal of information potentially relevant to the assessment of its contributions to its students' development of particular desired outcomes (e.g., GPAs, GREs, SATs, College BASE Examinations, Course and Instructor Critiques, et al.). Effective assessment demands that these data be integrated into meaningful conceptual wholes (i.e., comprehensive theoretical educational models). It is only within such contexts that information from various assessment activities can be meaningfully interpreted and effectively applied. In addition to raising the general level of awareness of these educational and assessment issues, the Educational Outcomes Assessment Working Group also demonstrated the capacity of an interdisciplinary faculty group to develop the collective talent to conduct valid and reliable educational assessment.

Phase II: New challenges, new opportunities

Despite the unique Academy context, the work of the group has been rewarded with invitations to speak and share techniques at national educational conferences sponsored by the American Association of Higher Education, North Central Association of Colleges and Universities, National Conference for Student Success and the College Reading and Learning Association Annual Conference. Additionally, group members were awarded Air Force Achievement Medals and equivalent civilian awards. As further evidence of the institutional value of this work, the Dean of the Faculty approved an ambitious charter designed to augment and extend the Phase I work and allocated the funds necessary to support its implementation. A period of negotiation among the Academy's Permanent Professors followed; resources and faculty hours committed to this project were reduced to 2 faculty man hours per department per week. Individual faculty members and cadets were also encouraged to volunteer to participate in assessment activities; however, participation rates varied considerably across the four academic divisions.

Under the new charter, the primary purpose of the Educational Outcomes Assessment Working Group was to continue to provide academic policy makers with the information needed to make curricular decisions and continuously improve the effectiveness of academic programs and practices. The group was tasked with the development and refinement of a variety of assessment tools (including performance evaluations, attitudinal surveys and other process metrics) to insure access to relevant, reliable and valid data concerning the educational outcomes and related academic processes. The group was also asked to develop supplementary assessment protocols and standards of assessment practice designed to increase understanding and enhance organizational trust. Increased decentralization, continued development and more inclusive involvement were to characterize Phase II activities and distinguish them from Phase I activities. Three particular educational outcomes were selected for assessment during Phase II. These were: 1) integrated fundamental knowledge; 2) the ability to frame and resolve ill-defined problems; and 3) intellectual curiosity. These three outcomes were selected because the first dealt with knowledge (which was initially assumed to be one of the easiest to measure objectively) and the

other two focused on the development of skills and attitudes and had been the subject of much of the developmental work accomplished during Phase I.

The ***Divisional Educational Outcomes Assessment Working Groups*** established in each of the four academic divisions at the conclusion of Phase I became a primary focus of assessment activities for Phase II. These groups, chaired by experienced individuals from the original Working Group designated by the senior Permanent Professor in each academic division, included at least one representative from each academic department. Their initial task was to review existing assessment data and conduct analyses to assess the contributions of each of the core (i.e., general education) courses offered within their respective academic divisions to each of the educational outcomes. Supplemental surveys, focus groups and other assessment activities were to be used to garner additional data. Reports from each divisional group were to be coordinated through the faculty working group and submitted to the Faculty Working Group by the end of the Spring Semester. An invaluable benefit of the initial assessment of the divisional core courses contributions to three of the educational outcomes was to provide training and experience for all new departmental assessment representatives.

As a complement to divisional assessment activities, ***Interdisciplinary "Horizontal" Assessment Teams*** were formed to examine contributions to the educational outcomes by all concurrent core courses. Three such groups were formed: One to examine freshman courses, another for sophomore courses and a third to consider junior and senior level core courses. The 30 courses which comprise the Air Force Academy's general education core include 6 to 8 courses from each of the four academic divisions (viz., Basic Sciences, Social Sciences, Humanities and Engineering). Each of these assessment teams was chaired by experienced members of the Phase I Educational Outcomes Assessment Working Group and included students and non-academic staff. These groups provided a "horizontal" look at cadets' educational development and allowed for convergent validation of the assessment of each of the individual courses to the three educational outcomes.

Another important point is that the teams agreed to use the 7-point rating scale listed below. Despite concerns about "rampant reductionism" and the inherent imprecision of assessment metrics, the advantages of a common rating scale seemed to be well understood by all the teams.

Course Contribution Rating Scale:

7 - Best possible contribution; reflects state of the art educational philosophy and practice; consistently solid implementation across all sections of the course.

6 - Excellent contribution; no deficiencies noted; solid consistency across sections; substantial evidence of appropriate policies and sound teaching practices.

5 - Slightly above average contribution; certain policies and practices may be particularly noteworthy but consistency across sections or course segments could improve.

4 - Satisfactory level of contribution; course does its part but impact on students' development of educational outcome is average.

3 - Slightly below average contribution; certain policies or practices appear inconsistent with sound educational practice; isolated evidence of negative impact on development of particular outcome.

2 - Poor contribution; evidence of general and/or continuing negative impact on development of particular outcome.

1 - Unsatisfactory level of contribution; course policy or procedures impede the development of desired educational outcome; general and continuing difficulties indicated.

Conclusion

Consistent with the goals of *assessment,* the activities undertaken by the Air Force Academy's Educational Outcomes Assessment Working Groups have been tailored to obtain the information needed to improve our ability to produce the outcomes we value. Assessment success requires the inclusion and cooperation of everyone from the Dean of Faculty to students themselves. A term which captures the uniqueness and emphasis of the Air Force Academy's approach to assessment is *integration.* Integration is the creation of wholes by bringing together diverse parts. The integration spanned across institutional agencies, educational philosophies, and experiences. The multi-faceted assessment plan incorporated both inprocess and output measures. Objective analysis of processes were continuously informed by and merged with subjective reactions of various organizational constituencies in order to sustain trust. Accomplishment of the Phase II has involved recruiting and training over 30 additional Assessment Working Group members from all academic divisions. This has distributed assessment expertise through approximately ten percent of the faculty and administration, and will help to insure the continued development, relevancy, and effectiveness of institutional assessment efforts[1].

[1]Note: Opinions expressed in this paper are those of the authors' alone and not necessarily the official policy of the United States Air Force Academy or any other government agency. Previous versions of this report were published in the North Central Association of Colleges and Schools' 1996 A Collection of Papers on Self-Study and Institutional Improvement and the Journal of Adult Assessment.

Chapter 2

Inputs and Instruments

CHAPTER TWO - INPUTS AND INSTRUMENTS

This chapter will provide a brief background concerning some of the instruments and inputs considered by the various assessment teams. The issue of inputs was discussed early and extensively by the faculty-wide Educational Outcomes Assessment Working Group. The general philosophy was to be as inclusive as possible; to provide as much relevant data for the teams to consider as was available. First, there was a review of available archival data (already gathered and on-file) and ways in which it might relate to the overall assessment charter. Second, the need for gathering new data was considered. There was a keen awareness of the overall constraint that departments would be tasked to provide no more than 2 hours per week to the assessment effort; this limited the number and type of additional data gathering instruments employed. However, all the teams also relied on many informal inputs, some of which are contained in their formal reports in the chapters which follow. Divisional Teams were comprised of representatives from each department; these representatives provided copies of course syllabi, examples of assignments and evaluations and also qualitative descriptions of course activities. Some of these materials were also passed to horizontal teams. Horizontal teams also gathered informal feedback from a variety of sources (cadets and junior faculty) to help them interpret apparent patterns in the data. Thus this chapter is designed to give the reader a flavor of the scope and variety of inputs considered by the teams. It is not a comprehensive account of all the inputs or information considered by all the teams.

This chapter discusses five different formal inputs: the College BASE; student course critiques, the course characteristics survey, the instructor outcomes emphasis survey and the student critical thinking survey. After a brief description, copies of the instruments themselves and tables with actual results will be presented.

The College BASE

The College BASE is a "criterion-referenced achievement test that assesses student proficiency in English, Mathematics, Science, Social Studies, and three cross disciplinary, cognitive competencies..." (College Basic Academic Subjects Examination, Guide to Test Content, p. 1). It was prepared at the University of Missouri-Columbia Center for Educational Assessment in 1989 and was specifically designed to address the content knowledge and thinking skills of students completing general education curricula. Its knowledge section is organized hierarchically at four distinct levels: subjects, clusters, skills and subskills.

This instrument was administered in 1989 and its results had already been widely circulated. However, the working group agreed that such a nationally-normed knowledge measure might provide a useful general context for looking at the "integrated knowledge" outcome in particular. In addition to providing national norms, the test's internal consistency (i.e., a "High" in Math meant about the same as a "High" in English) would allow assessors to make broad general comparisons of net gains in different areas.

The College BASE test structure:

Subjects	Clusters	Skills
English	Reading & Literature	*Reading Critically, Reading Analytically, Understanding Literature*
	Writing	*Writing as a Process, Writing Conventions, Writing Exercise*
Mathematics	General Math	*Practical Applications, Properties & Notations, Using Statistics*
	Algebra	*Evaluating Expressions, Equations and Inequalities*
	Geometry	*2&3 Dimensional Figures, Geometric Calculations*
Science	Laboratory & Field Work	*Observation/Experimental Design, Techniques, Interpreting Results*
	Fndmntl Concepts	Physical Sciences, *Life Sciences*
Social Studies	History	*Significance of World Events, Sig. of U.S. Events*
	Social Sciences	*Geography, Political/Economic Structures, Social Science Procedures*

The entire College BASE test would take approximately 10 hours for a single student to complete. However, each complete test can be broken into 10 distinct sections which can be accomplished simultaneously by 10 different students. This approach was used at the Air Force Academy when the test was administered in both 1989 and 1993. On both occasions random and representative samples of freshmen (who had not completed any of the USAFA core) and seniors (who had completed 90% of their core requirements) completed the examination. About 250 students in each group (freshmen and seniors) participated. Results are reported as the proportion of students scoring in each of three competency categories: High, Medium and Low. The test also contained a subsection on thinking skills. The working group found the College BASE's thinking skills to be very comparable to the rubric developed for assessing cadets' ability to frame and resolve ill-defined problems. In fact, the match between what the College BASE referred to as *Adaptive Reasoning* and what the group had defined as *Excellent Ability to Frame and Resolve Ill-Defined Problems* was remarkably similar. Attempts to locate the original 1993 College BASE report were unsuccessful; however, summary data of overall scores from 1993 and 1989 revealed that the two administrations were nearly identical for each of the four primary knowledge categories (English, Math, Science and Social Studies) as well as the overall score. Results from the appraisal of cadets' thinking abilities also converged with the data derived from a comprehensive assessment of senior cadets in 1995 (Phase I: Final Report). On average, about 10% more seniors achieved "High" scores than did freshmen. Complete results are listed on the following pages. Although not relied upon heavily by individual assessment teams, program implications for core course contributions to integrated knowledge and thinking ability are discussed later.

One department head argued vigorously that the College BASE are very limited in the amount of actual information they provide due to their relative ease and simplicity. Clearly scores that place the institution at the 99th percentile nationally support such skepticism regarding the examination's assessment suitability. However, as a bench mark for the Academy's general "literate citizen" criteria, the College BASE has certain advantages. It is nationally normed for students both starting and completing general education (i.e., "core") requirements. It also allows comparisons between many diverse disciplines covered by core courses at the Academy. It even provides some general comparisons of the relative starting points and changes attributable to divisional core programs (i.e., sequences within the four respective academic divisions).

INSTITUTIONAL SUMMARY REPORT

COLLEGE BASE

COLLEGE BASIC ACADEMIC SUBJECTS EXAMINATION

Institution: United States Air Force Acad.
Location: US Air Force Acad, CO 80840
FICE No.: 001369
Test Date: SEPTEMBER 1990
Test Form: XA
Number of complete tests: 38
Number of incomplete tests: 0

Freshmen

*** MATRIX TEST FORM ***
NUMBER OF ANSHER SHEETS: 255

THE RIVERSIDE PUBLISHING COMPANY

This INSTITUTIONAL SUMMARY REPORT presents aggregate scores for all examinees who took *College BASE* on the test date noted above and who indicated enrollment at your institution. Scores on *College BASE* are presented in two dimensions of achievement: four Subject scores (English, mathematics, science, and social studies) and three Competency scores (interpretive, strategic, and adaptive reasoning). Subject scores show examinees' familiarity with traditional subject matter. For diagnostic purposes, the Subjects are divided into levels that become increasingly more specific: from Subjects to Clusters to Skills. Subject scores, and the associated Cluster scores, are reported on a scale ranging from 40 to 560, with an average being 300. Skill scores are reported as High, Medium, or Low.

Competency scores show examinees' capability in cognitive processing skills as they apply to all subjects, regardless of particular content. These scores are reported as High, Medium, or Low. The Composite score (representing overall performance on the examination) is reported on the same 40-to-560 scale as the Subject scores. For each *College BASE* scaled score, this INSTITUTIONAL SUMMARY REPORT displays the average score (representing the median) and a standard deviation (representing the range encompassing approximately 68 percent of the examinees' scores). For the *College BASE* scores that are designated as High, Medium, or Low, this INSTITUTIONAL SUMMARY REPORT gives the number (and the corresponding percentage) of examinees who achieved a High rating (H), a Medium rating (M), and a Low rating (L). Please refer to your *Institutional Summary Report: Interpretive Guide* for more complete information.

SUBJECT SCORES

ENGLISH

Avg.	S.D.
323	28

q5

READING & LITERATURE		reading critically			reading analytically			understanding literature		
Avg.	S.D.	H	M	L	H	M	L	H	M	L
326	31	10	27	1	13	21	4	8	29	1
		26%	71%	3%	34%	55%	11%	21%	76%	3%

WRITING		writing as a process			conventions of written English			writing exercise		
Avg.	S.D.	H	M	L	H	M	L	H	M	L
306	31	12	23	3	8	28	2	0	33	5
		32%	61%	8%	21%	74%	5%	0%	87%	13%

MATHEMATICS

Avg.	S.D.
382	30

q9

GENERAL MATHEMATICS		practical applications			properties and notations			using statistics		
Avg.	S.D.	H	M	L	H	M	L	H	M	L
360	37	16	22	0	23	15	0	17	21	0
		42%	58%	0%	61%	39%	0%	45%	55%	0%

ALGEBRA		evaluating expressions			equations & inequalities		
Avg.	S.D.	H	M	L	H	M	L
358	27	13	25	0	23	15	0
		34%	66%	0%	61%	39%	0%

GEOMETRY		2- & 3-dimensional figures			geometrical calculations		
Avg.	S.D.	H	M	L	H	M	L
376	25	17	21	0	22	16	0
		45%	55%	0%	58%	42%	0%

SCIENCE

Avg.	S.D.
361	28

q5

LABORATORY & FIELD WORK		observation/experimental design			laboratory/field techniques			interpreting results		
Avg.	S.D.	H	M	L	H	M	L	H	M	L
361	29	18	20	0	20	18	0	14	24	0
		47%	53%	0%	53%	47%	0%	37%	63%	0%

FUNDAMENTAL CONCEPTS		life sciences			physical sciences		
Avg.	S.D.	H	M	L	H	M	L
336	28	7	27	4	20	18	0
		18%	71%	11%	53%	47%	0%

SOCIAL STUDIES

Avg.	S.D.
341	25

q6

HISTORY		significance of world events			significance of U.S. events		
Avg.	S.D.	H	M	L	H	M	L
336	29	14	24	0	4	33	1
		37%	63%	0%	11%	87%	3%

SOCIAL SCIENCES		geography			political/economic structures			social science procedures		
Avg.	S.D.	H	M	L	H	M	L	H	M	L
346	31	25	13	0	16	21	1	13	23	2
		66%	34%	0%	42%	55%	3%	34%	61%	5%

COMPETENCY SCORES

INTERPRETIVE REASONING			STRATEGIC REASONING			ADAPTIVE REASONING		
H	M	L	H	M	L	H	M	L
37	1	0	25	13	0	4	34	0
97%	3%	0%	66%	34%	0%	11%	89%	0%

COMPOSITE SCORE

355

This score is the average composite score achieved by this group of examinees.

INSTITUTIONAL SUMMARY REPORT

COLLEGE BASE

COLLEGE BASIC ACADEMIC SUBJECTS EXAMINATION

Institution: United States Air Force Acad.
Location: US Air Force Acad, CO 80840
FICE No.: 001369
Test Date: AUGUST 1990
Test Form: XA
Number of complete tests: 40
Number of incomplete tests: 0

Seniors

*** MATRIX TEST FORM ***
NUMBER OF ANSWER SHEETS: 245

THE RIVERSIDE PUBLISHING COMPANY

This INSTITUTIONAL SUMMARY REPORT presents aggregate scores for all examinees who took *College BASE* on the test date noted above and who indicated enrollment at your institution. Scores on *College BASE* are presented in two dimensions of achievement: four Subject scores (English, mathematics, science, and social studies) and three Competency scores (interpretive, strategic, and adaptive reasoning). Subject scores show examinees' familiarity with traditional subject matter. For diagnostic purposes, the Subjects are divided into levels that become increasingly more specific: from Subjects to Clusters to Skills. Subject scores, and the associated Cluster scores, are reported on a scale ranging from 40 to 560, with an average being 300. Skill scores are reported as High, Medium, or Low.

Competency scores show examinees' capability in cognitive processing skills as they apply to all subjects, regardless of particular content. These scores are reported as High, Medium, or Low. The Composite score (representing overall performance on the examination) is reported on the same 40-to-560 scale as the Subject scores. For each *College BASE* scaled score, this INSTITUTIONAL SUMMARY REPORT displays the average score (representing the median) and a standard deviation (representing the range encompassing approximately 68 percent of the examinees' scores). For the *College BASE* scores that are designated as High, Medium, or Low, this INSTITUTIONAL SUMMARY REPORT gives the number (and the corresponding percentage) of examinees who achieved a High rating (H), a Medium rating (M), and a Low rating (L). Please refer to your *Institutional Summary Report: Interpretive Guide* for more complete information.

SUBJECT SCORES

ENGLISH

Avg.	S.D.
346	28

97

READING & LITERATURE Avg.	S.D.
338	35

reading critically	H	M	L
	18	21	1
	45%	52%	2%

reading analytically	H	M	L
	18	21	1
	45%	52%	2%

understanding literature	H	M	L
	16	24	0
	40%	60%	0%

WRITING Avg.	S.D.
333	31

writing as a process	H	M	L
	22	17	1
	55%	42%	2%

conventions of written English	H	M	L
	8	32	0
	20%	80%	0%

writing exercise	H	M	L
	0	38	2
	0%	95%	5%

MATHEMATICS

Avg.	S.D.
406	29

99

GENERAL MATHEMATICS Avg.	S.D.
386	30

practical applications	H	M	L
	31	9	0
	77%	22%	0%

properties and notations	H	M	L
	13	27	0
	32%	67%	0%

using statistics	H	M	L
	26	14	0
	65%	35%	0%

ALGEBRA Avg.	S.D.
373	26

evaluating expressions	H	M	L
	24	16	0
	60%	40%	0%

equations & inequalities	H	M	L
	24	16	0
	60%	40%	0%

GEOMETRY Avg.	S.D.
380	22

2- & 3-dimensional figures	H	M	L
	19	21	0
	47%	52%	0%

geometrical calculations	H	M	L
	27	13	0
	67%	32%	0%

SCIENCE

Avg.	S.D.
381	39

99

LABORATORY & FIELD WORK Avg.	S.D.
375	35

observation/experimental design	H	M	L
	11	28	1
	27%	70%	2%

laboratory/field techniques	H	M	L
	32	8	0
	80%	20%	0%

interpreting results	H	M	L
	21	19	0
	52%	47%	0%

FUNDAMENTAL CONCEPTS Avg.	S.D.
361	27

life sciences	H	M	L
	25	15	0
	63%	38%	0%

physical sciences	H	M	L
	26	14	0
	65%	35%	0%

SOCIAL STUDIES

Avg.	S.D.
367	25

99

HISTORY Avg.	S.D.
353	29

significance of world events	H	M	L
	21	19	0
	52%	47%	0%

significance of U.S. events	H	M	L
	16	24	0
	40%	60%	0%

SOCIAL SCIENCES Avg.	S.D.
368	29

geography	H	M	L
	28	12	0
	70%	30%	0%

political/economic structures	H	M	L
	28	11	1
	70%	27%	2%

social science procedures	H	M	L
	18	22	0
	45%	55%	0%

COMPETENCY SCORES

INTERPRETIVE REASONING	H	M	L
	39	1	0
	97%	2%	0%

STRATEGIC REASONING	H	M	L
	39	1	0
	97%	2%	0%

ADAPTIVE REASONING	H	M	L
	9	30	1
	22%	75%	2%

COMPOSITE SCORE

97th %

380

This score is the average composite score achieved by this group of examinees.

Student Course Critiques

The Air Force Academy began using institutionally unique course critiques late in the last decade. Several iterations have helped streamline these forms. Because of the Academy's extremely high faculty turnover (approximately 25% per year) and low teaching experience (about 2 years on average), these instruments have come to play a very significant role in both departmental and faculty-wide faculty development efforts. Although some residual skepticism remains in some departments, there seems to be a general acceptance and appreciation of what the educational literature suggests: although imperfect, students' perceptions (as captured by end of course critiques) are relevant to the attainment of educational outcomes. Overall the correlation between critique ratings and independent measures of merit appears to be about .4 (about the same as the relationship between IQ scores and GPA at most schools). The Academy found that in the first two years of adopting a comprehensive critique system, overall average ratings increased by nearly a full standard deviation (viz., the "average" USAF Academy faculty member in 1991 received scores that would have place him/her on the 80th percentile in 1989).

Student opinions are not equally valid with respect to all three of the educational outcomes addressed in this assessment. In fact, many core courses are introductory in nature; they provide students with their first formal training in subjects such as political science, ethics or thermodynamics. Students may be unable to evaluate or to accurately report the contribution the course has truly made to their level of integrated knowledge about such subjects. Objective tests are much better suited for this purpose. However, students are very capable of reporting the effects of classroom experiences on their attitudes. The validity of students' self assessment of their own skills such as the ability to frame and resolve ill-defined problems seems to fall somewhere in between the validity of their self-assessments of their knowledge and attitudes. After considering summary data from the last several semesters, the working group selected five items which appeared to have logical connections to the three outcomes. These items (and the outcome to which they were assumed to most closely relate) are listed below.

(integrated knowledge)
#21. Relevance and usefulness of course content was:
#22. Amount you learned in this course was:

(ill-defined problem resolution)
#16. Intellectual challenge and encouragement of independent thought were:
#36. This course improved my ability to deal with problems which don't have "approved solutions".

(intellectual curiosity)
#39. There are a number of things in this general subject area I'd like to read more about.

Items 1-23 on the critique were answered with respect to a six point Likert scale ranging from: very poor (1); poor (2); fair (3); good (4); very good (5) and excellent (6). Items 23-40 on the critiques were also answered with respect to a six point Likert scale, but the six alternatives were slightly different: strongly disagree (1); disagree (2); slightly disagree (3); slightly agree (4); agree (5) and strongly agree (6). Averages shown in the following table reflect the grand course averages for all listed courses from the Spring and Fall semesters of 1995.

Course	Independent Thought	Relevance of Content	Amount Learned	Problems w/o Approved Solutions	Read More
DFAN 110	4.64	4.31	3.97	5.47	5.11
DFAN 215	4.43	4.47	4.46	4.85	5.09
DFAN 310	4.13	3.61	3.94	4.57	4.14
DFAS 320	4.36	4.00	4.38	4.94	5.24
DFAS 410	3.96	3.45	3.88	4.43	4.47
DFB 215	4.44	4.41	4.41	4.48	4.97
DFBL 110	4.79	4.61	4.63	5.36	5.32
DFBL 310	4.79	5.16	4.90	5.76	5.57
DFC 141	4.93	4.61	4.83	5.07	4.76
DFC 142	4.39	4.18	4.27	4.86	4.67
DFCE 310	3.54	3.26	3.65	4.75	3.73
DFCS 110	4.41	3.76	4.18	4.94	4.35
DFEE 215	4.39	4.06	4.43	4.57	4.37
DFEE 225	4.83	4.35	4.62	5.02	4.79
DFEE 231	4.55	4.28	4.45	4.87	4.87
DFEE 311	4.59	4.13	4.34	5.15	4.75
DFEG 221	4.39	4.29	4.23	5.00	4.65
DFEG 310	4.70	4.66	4.58	5.30	5.24
DFEM 120	4.35	3.99	4.14	4.68	4.06
DFENG 111	4.49	4.07	3.98	4.86	4.42
DFENG 211	4.64	3.94	4.04	4.93	4.65
DFENG 311	4.30	4.16	3.81	4.79	4.26
DFF 131	4.73	4.78	4.82	5.01	5.61
DFF 132	4.51	4.50	4.59	4.93	5.06
DFF 141	4.73	4.73	4.59	5.06	5.33
DFF 142	4.55	4.50	4.32	4.82	4.81
DFF 150	4.91	4.83	4.52	5.39	5.47
DFH 101	4.49	4.39	4.32	4.96	5.39
DFH 202	4.82	4.86	4.91	5.38	5.68
DFH 374	5.43	5.29	5.29	5.71	5.57
DFL 320	4.70	5.03	4.95	5.46	5.50
DFL 420	4.98	5.50	5.21	5.96	5.81
DFM 210	4.40	4.64	4.39	5.26	5.22
DFMS 141	4.58	4.48	4.37	4.98	4.44
DFMS 142	4.66	4.25	4.41	4.57	4.10
DFMS 152	4.51	4.35	4.06	4.98	4.53
DFMS 220	4.17	3.90	3.91	4.61	3.83
DFMS 358	5.14	4.94	5.14	5.31	4.94
DFP 110	4.79	4.62	4.56	5.31	5.15
DFP 215	4.69	4.24	4.41	5.00	4.64
DFPFA 310	4.58	4.29	4.12	5.32	4.62
DFPS 211	4.71	4.87	4.65	5.24	5.30
DFPS 212	4.61	4.72	4.58	5.37	5.15
ENGR 410	3.54	3.10	2.95	4.77	3.58
MAS 220	4.18	4.49	4.34	4.99	4.91
MAS 330	4.07	4.24	4.06	5.09	4.96
MAS 440	4.27	4.57	4.40	5.55	6.00
Avgerage	4.58	4.46	4.44	5.07	4.90

Course Critique: Two Semester Averages

Course Characteristics Survey

Several members of the working group had previously worked on the Academy's initial effort to develop a baseline by conducting a sample survey of curricular characteristics. This initial curriculum inventory effort (EOAWG Phase I: Initial Report) had encountered some resistance. Questions were automatically assumed to be evaluative in effort and there was readily apparent discomfort shown by some faculty members at not knowing what the "approved" solution might be, as well as concern about how these data might be used in future. It turned out to be much more difficult than anticipated to develop an instrument which would show meaningful variation in courses without creating undo reactance or suspicion. The group's experience from the previous two years guided the development of the instrument used to collect course characteristic data for this report.

When the instrument was complete, divisional assessment chairs were asked to train departmental representatives to individually sit down and explain the form and the items it contained to each of the respective core course directors in their department, then return the numbers for inclusion in a common data base. Apparent misunderstandings, inconsistencies or questions were handled through departmental representatives. This mediated form of data collection through "trusted agents" who were familiar with the purpose of the project was one of the lessons learned from the previous course characteristics data gathering efforts.

Page 20 contains a blank form and the following page an actual example of data from an individual core course, in this case Leadership. The far left column contains a number events from which a student's grade in a particular course might be determined. The first data column asks for the raw number, but the following column asks for the percentage of the overall grade allocated to this activity in the course; this column should add up to 100%. The following 4 columns (designated by small arrows at the top) are meant to pose four exhaustive, more specific categories of graded activity and relate to the general activity in the left column. So, using the Leadership course as an example: the bottom line tells us that 10 events which account for 20 percent of students total grade fall in the category of IP (Instructor Prerogative) Points. Across the 25 or so sections about half these points are earned by students answering multiple choice questions, a quarter by responses to essay questions and the final quarter by student performances (speeches, debates, skits, etc.).

Summary data suggest that nearly all core courses involve final examinations (which are slightly more than 25% of the course grade); on average, slightly more than 2 GRs (Graded Reviews) are also given which account for another 31% of students grades. The proportion of particular graded activities are: 27% multiple choice; 36% essay; 26% computation; and 12% performance. Computation was defined as a problem that required some mathematical formula to arrive at a particular and precise "approved solution".

The next column asks whether each graded activity occurred in or out of class. The final three columns relate to three pedagogical alternatives suggested by the literature, and previous Educational Outcomes Assessment Group studies, to relate to the educational outcomes under study. The Air Force Academy's strict Honor Code requires careful attention to the instructions students are given concerning the extent to which they may interact with classmates in completing

graded assignments. At the Academy, complete individual effort (isolation from interaction) is the default position. Permission to collaborate or work with other students must be explicitly stated and is tightly controlled. Scores shown in the summary statistics on the third page are actually the arithmetic compliments (100-x) of the description on the work sheet. Thus, the overall average of 13.6% represents the extent to which collaborative activity is encouraged in core courses. In fact, the relatively large standard deviation (17.9) reflects a marked diversity of opinion and philosophy concerning group work. Over a third of the core courses allow nearly no interactive effort (5% or less collaboration) while one course requires 95% collaboration and another four require a substantial amount.

The next column shows the amount of task reaccomplishment allowed. Students often learn from making mistakes. If a course places a high penalty on student mistakes, there is likely to be less active exploration. One way to encourage such exploration is to allow students to recover from mistakes. The three categories reflected in the "Reaccomplish" column show the extent to which students were allowed to resubmit or reaccomplish work. Examples of reaccomplishment with **P**artial compensation would be the opportunity to resubmit an unsatisfactory or poor paper for partial credit or to reaccomplish an examination with a group of students to earn bonus points. Examples of **F**ull credit would be the opportunity to reaccomplish either tests or projects for full credit (i.e., all the points allotted to the activity). Average scores for each course were derived by computing a weighted average of the rules for all graded events in the course. As can be seen by the overall average of 7.6%, the idea of allowing students to reaccomplish some portion of their work is not popular among faculty. In fact, the vast majority of courses limit reaccomplishment opportunities to less than 10% of the total grade.

The final column relates to the degree of choice afforded to cadets. Such choice might include the opportunity to choose between taking quizzes or doing projects for extra credit or might relate to simply being given the option to define your own project or paper. Once again the results of the core course curricular survey suggested that only a relatively small fraction of the graded activities in core courses involve cadet choice. It is also again apparent that several course courses actively incorporate student choices within their curricular requirements, but the majority of core courses provide few choices to students. It is worth reiterating that these results were derived from one-on-one sessions with respective core course directors; anomalies were clarified and outliers reconfirmed. Even without subsequent analyses, these data could provide the basis for discussion and increased understanding of curricular and pedagogical alternatives among the various core courses.

Course _____ Director _____ Phone _____

Graded Events
(average section)

These four should add up to 100

↓ ↓ ↓ ↓

	#	% Total Grade	% Multiple Choice	% Essay	% Computation	% Performance	In/Out Class (I) or (O)	Honor ROE % Ind Effort 0 - 100%	Reaccomplish: No,Partial or Full Credit (N)(P) or (F)	% Options Choice
Final										
GRs										
Quizzes										
Laboratories										
Homework										
Projects										
Others (describe) _____ _____ _____										
IP Points										

Course Beh Sci 310 Leadership Director Hemdr Botonis Phone 3860

Graded Events
(average section)

These four should add up to 100

	#	% Total Grade	% Multiple Choice	% Essay	% Computation (← carefully w/ approved solution)	% Performance (build, make, do, contribute, participate etc.)	In/Out Class (I) or (O)	Honor ROE % Ind Effort 0 - 100% ("Individual Effort" will be 100%. Group project w/ no individual separate grade worth 0%.)	Reaccomplish: No, Partial or Full Credit (N)(P) or (F) (After required completion. Don't count "encouraging early drafts". Full credit would have been available if project was first time submitted.)	% Options/ Choice (- Writing a paper/essay on anything you want would be 100%. Choosing 1 of 2 possible essay/projects would be 50% or all multiple choice would be 0% project % choice. 0% choice.)
Final	1	30	75	25	—	—	I	100	N	0%
GRs	2	30	100	—	—	—	I	85	P	0%
Quizzes	Ø	—								
Laboratories	Ø	—								
Homework	Ø	—								
Projects	10	20	—	100	—	—	O	85	F	50%
Others (describe) i.e. journals, skits, leading class discussion, production, etc.	Ø	—								
IP Points	10	20	50	25	—	25	I	85	P	20%

a:\ccourse.sav

	division	course	final grade % fnlgr	number of grs grnum	GR % fnl grgr	multiple choice % fnl/grd pmc	essay % fnl pes	computations % fnl pcmp	Assignments % fnl pprf	out-of-class % fnl poc	collaboration on test pgrp	reaccomplishment permitted reacc	on-test segment optional popt
1	ss	BSci110	27.0	2	31.0	40.0	40.0	3.0	17.0	25.0	17.0	43.0	34.0
2	ss	BSci310	30.0	2	30.0	62.5	32.5	.0	5.0	20.0	11.0	45.0	14.0
3	ss	Econ221	30.0	2	45.0	50.0	44.0	.0	6.0	17.5	2.0	.0	.0
4	ss	Econ310	35.0	2	45.0	49.0	42.0	.0	9.0	10.0	2.0	.0	.0
5	ss	Law320	30.5	2	28.0	43.0	46.0	.0	10.0	10.0	10.0	10.0	.0
6	ss	Law420	31.3	2	50.0	26.0	64.0	.0	10.0	6.0	.0	.0	.0
7	ss	Mgt210	25.0	2	20.0	50.0	23.0	.0	27.0	50.0	35.0	.0	19.0
8	ss	PSci211	25.0	2	30.0	44.0	50.0	.0	6.0	35.0	10.0	.0	.0
9	ss	PSci212	25.0	2	30.0	44.0	51.0	.0	5.0	35.0	10.0	.0	.0
10	hm	Eng111	.0	0	.0	.0	75.0	.0	25.0	75.0	.0	7.5	60.0
11	hm	Eng211	25.0	1	15.0	9.0	81.0	.0	10.0	10.0	3.0	.0	45.0
12	hm	Eng311	20.0	0	.0	4.0	51.0	.0	45.0	35.0	11.0	.0	78.0
13	hm	FLf142	23.0	2	29.0	11.0	70.0	.0	19.0	17.0	24.0	20.0	22.0
14	hm	FLc132	20.0	2	25.0	32.0	10.0	.0	58.0	8.0	.0	4.0	1.0
15	hm	Hist101	25.0	2	20.0	26.0	69.0	.0	5.0	30.0	.0	3.0	62.0
16	hm	Hist202	25.0	2	30.0	43.0	53.0	.0	5.0	25.0	.0	3.0	58.0
17	hm	Phil310	30.0	1	15.0	6.0	86.0	.0	8.0	35.0	8.0	.0	.0
18	bs	bio215	25.0	3	37.5	41.0	57.0	2.0	.0	19.0	27.0	.0	10.0
19	bs	Chm141	30.3	3	45.4	.0	20.0	80.0	.0	9.0	.0	.0	.0
20	bs	Chm142	27.0	3	40.6	3.0	18.0	79.0	.0	14.0	.0	.0	.0
21	bs	CSci110	25.0	2	30.0	22.0	30.0	2.0	46.0	30.0	8.0	14.0	6.0
22	bs	Math141	26.0	3	47.0	38.0	21.0	41.0	.0	15.0	19.0	.0	.0
23	bs	Math142	25.0	3	55.0	26.0	9.0	65.0	.0	12.0	7.0	3.0	1.0
24	bs	Math220	30.0	3	45.0	40.0	19.0	39.0	2.0	12.0	6.0	.0	.0
25	bs	Phys110	29.7	3	41.7	30.0	18.0	48.0	2.0	13.0	18.0	6.0	5.0
26	bs	Phys110z	37.0	3	26.0	50.0	6.0	36.0	8.0	14.0	29.0	15.0	.0
27	bs	Phys110h	31.0	3	37.0	28.0	14.0	52.0	6.0	8.0	17.0	.0	4.0
28	bs	Phys215	28.0	3	42.0	35.0	18.0	44.0	3.0	5.0	5.0	5.0	5.0
29	bs	Phys215h	26.4	3	39.6	30.0	23.0	36.0	10.0	15.0	.0	.0	.0
30	eg	Aero215	27.8	1	16.7	24.0	18.0	52.0	6.0	43.0	13.0	7.0	5.0
31	eg	Ast320	35.0	1	25.0	.0	31.0	69.0	.0	40.0	8.0	19.0	7.0
32	eg	Ast410	35.0	1	25.0	.0	45.0	47.0	8.0	30.0	15.0	.0	12.0
33	eg	Eng310	25.0	2	47.0	38.0	8.0	54.0	.0	28.0	23.0	10.0	8.0
34	eg	Eng410	10.0	1	5.0	5.0	16.0	10.0	69.0	26.0	95.0	26.0	.0
35	eg	EEng215	25.0	4	48.0	.0	22.0	73.0	5.0	.0	8.0	3.0	.0
36	eg	EEng231	25.0	3	33.0	6.0	11.0	78.0	5.0	16.0	.0	.0	4.0
37	eg	EngM120	27.6	3	41.4	18.0	7.0	64.0	12.0	5.0	21.0	15.0	.0
38	eg	CEng310	.0	2	24.0	36.0	54.0	5.0	5.0	30.0	54.0	30.0	.0
$\bar{X}$			25.7	2.1	31.4	26.6	35.6	25.8	12.0	21.8	13.6	7.6	12.1
SD			(7.8)	(.9)	(13.6)	(18.2)	(22.8)	(29.7)	(16.4)	(15.0)	(17.9)	(11.7)	(20.8)

Instructor Survey

Instructor attitudes and opinions are strongly related to student outcomes. The self-fulfilling prophecy is perhaps one of the most robust phenomena in social psychology. It also one of the most pertinent to education. The Faculty Survey section of the Phase I: Initial Report represented the beginning of our attempts to quantify the differences (and similarities) among USAFA faculty. Relying on an anonymous computer response format, initial results showed that there was considerable variation among faculty concerning their attitudes and opinions as well as their demographic characteristics. Results from this survey were then grouped by academic department and compared to other measures such as grades given to students and students' ratings of course and instructor contributions to their learning. No significant relationships between faculty attitudes and grades given were found. However, substantial relations between faculty attitudes and students' ratings of departmental course contributions to educational outcomes were apparent in nearly every instance (Educational Outcomes Assessment Working Group, Phase I: Final Report, pp. 37-48). An additional study on a smaller (but still largely representative) sample of faculty members contained in the same report suggested that faculty members view of education (viz. their attitudes relating to the relative importance of cadet attitudes) was the single most significant predictor of their subsequent course critique ratings. At an individual level, some psychological reciprocity and circularity is unavoidable: good ratings lead teachers to value student attitudes and focus upon them in the classroom. This focus, in turn, has facilitative effects on the classroom climate and student satisfaction and performance. However, at the broader level, such individual effects can be amplified even further. Directors of core courses exert a very strong influence on the numerous junior faculty members who teach for them. Discounting the potential importance of cadet attitudes (viz., their intellectual curiosity) can have a chilling effect on classroom climate. The assessment group recognized the importance of faculty attitudes and set out to collect relevant data. The following pages contain the summary report with attachments provided to each of the teams.

Procedural inconsistencies (i.e., core course instructor sample sizes varied from 1 to 19) and confusion about instructions (dealing with the interpretation of the % questions in the final 5 columns of the worksheet) served to limit the clarity of these results. The outcomes of particular interest in this study are numbers 1, 2 and 6 (integrated knowledge, ill-defined problems and intellectual curiosity respectively). There is wide disparity in the relative emphasis (viz. appropriate coverage) reflected by these results with respect to courses, departments and even divisions, while everyone seems to agree that fundamental knowledge deserves the heaviest coverage 2.53 (.61) on a three point scale. The amount of variance in response to the ill-defined problem outcome (.90) is the greatest of all the outcomes and the average weight (1.96) near the bottom. The intellectual curiosity outcome received a similar slightly below average coverage rating (1.93) but slightly lower standard deviation (.84). It was difficult for some assessment teams to award high scores to core courses where the course director or faculty felt that the level of coverage most appropriate for that course respect to a given outcome was only *light* (1.0) or even less.

MEMORANDUM FOR: DF Assessors 16 May 1996

SUBJECT: Instructor Assessment of Core Courses

The attached data comes from instructor and course director ratings. Inputs from Econ 310 are not included, but a new version of this data will be distributed as soon as that data is collected and tabulated. The organization of this package is as follows:

Attachment 1 **CD/Instructor Course Assessment Form.** This was the collection device for the data that follows. We deemed a portion of the data as "unusable" because instructors used different methods for filling-in the *percentage* measures. Thus, we only used the Coverage Scale (0-3) inputs.

Attachment 2 **Number of respondents by course.** Shows the number of non-blank responses for each Course x Educational Outcome.

Attachment 3 **Average Rating by Course.** Shows the average rating for each Course x Educational Outcome.

Attachment 4 **Average Rating by Department.** Shows the average rating for each Department x Educational Outcome

Attachment 5 **Average Rating by Division.** Shows the average rating for each Division x Educational Outcome.

Attachment 6 **Overall Average Rating for DF.** Shows the DF-wide average for each of the educational outcomes

Attachment 7 **Graphs of Division Output.** Two graphs: the first graph groups all seven outcomes by division, while the second graph groups the division average for each of the seven outcomes.

Attachment 8 **Graphs of Horizontal Output.** Taking horizontal cuts through the data allows one to see the trend from 100 level courses through 400 level courses to increased the emphasis on some outcomes while decreasing the emphasis on others. Two graphs are included: the first graph groups all seven outcomes by level, while the second graph groups the course levels for each of the seven outcomes.

As noted above, revisions to the attached data and graphical output will be revised and distributed when all the remaining data has been collected. There may be some changes to the DF-wide averages, but they should be relatively small. The only divisional averages that will change will be the Social Sciences Division.

Should you have any questions about the data or how to interpret the attachments, please call me at x3122 or send an email to armacostap@dfm.

ANDREW P. ARMACOST, Capt, USAF
Social Science Divisional Assessment "Stand-In"

Attachment 1 (p. 1 of 1)

*Course:*_____
*POC:*_____
*Phone:*_____

Outcomes	*Coverage Scale* None	Light	Avg.	Heavy	% Student Effort	%Time Out of Class	%Time in Class	% Points	Activities (Use Back)
1 Fundamental Knowledge	0	1	2	3					
2 Ill-Defined Problems	0	1	2	3					
3 Communicate Effectively	0	1	2	3					
4 Independent Learners	0	1	2	3					
5 Work With Others	0	1	2	3					
6 Intellectually Curious	0	1	2	3					
7 Military Professionals	0	1	2	3					

Number of Responses for Each Course

	OUTCOME						
	1	**2**	**3**	**4**	**5**	**6**	**7**
AEROENGR 215	4	4	4	4	4	4	4
ASTRO 320	5	5	5	5	5	5	5
ASTRO 410	7	7	7	7	7	7	7
BEH SCI 110	6	6	6	6	6	6	6
BEH SCI 310	8	8	8	8	8	8	8
BIOLOGY 215	7	6	6	6	6	5	6
CHEM 141	15	15	15	15	15	15	15
CHEM 142	19	19	19	19	19	19	19
CIV ENGR 310	10	10	10	10	10	10	10
COMP SCI 110	6	6	6	6	6	6	6
ECON 221	5	5	5	5	5	5	5
EL ENGR 215	5	5	5	5	5	5	5
EL ENGR 231	2	2	2	2	2	2	2
ENGLISH 111	11	11	11	11	11	11	11
ENGLISH 211	14	14	14	14	14	14	14
ENGLISH 311	5	5	5	5	5	5	5
ENGR 110	1	1	1	1	1	1	1
ENGR 310	5	5	5	5	5	5	5
ENGR 410	13	13	13	13	13	12	12
ENGRMECH 120	4	4	4	4	4	4	4
FRENCH 142	4	4	4	4	4	4	4
HISTORY 101	1	1	1	1	1	1	1
HISTORY 202	1	1	1	1	1	1	1
LAW 320	3	3	3	3	3	3	3
LAW 420	6	6	6	6	6	6	6
MATH 141	7	7	7	7	7	7	7
MATH 142	10	10	10	10	10	10	10
MATH 220	5	5	5	5	5	5	5
MGT 210	1	1	1	1	1	1	1
PHILOS 310	7	7	7	7	7	7	7
PHYSICS 110	10	10	10	10	10	10	10
PHYSICS 215	8	8	8	8	8	8	8
POL SCI 211	2	2	2	2	2	2	2

Attachment 2 (p. 2 of 3)

Number of Responses for Each Department

	OUTCOME						
	1	**2**	**3**	**4**	**5**	**6**	**7**
AEROENG	4	4	4	4	4	4	4
ASTRO	12	12	12	12	12	12	12
BEH SCI	14	14	14	14	14	14	14
BIOLOGY	7	6	6	6	6	5	6
CHEM	34	34	34	34	34	34	34
CIV ENGR	10	10	10	10	10	10	10
COMP SCI	6	6	6	6	6	6	6
ECON	5	5	5	5	5	5	5
EL ENGR	7	7	7	7	7	7	7
ENGLISH	30	30	30	30	30	30	30
ENGR	19	19	19	19	19	18	18
ENGRMECH	4	4	4	4	4	4	4
FRENCH	4	4	4	4	4	4	4
HISTORY	2	2	2	2	2	2	2
LAW	9	9	9	9	9	9	9
MATH	22	22	22	22	22	22	22
MGT	1	1	1	1	1	1	1
PHILOS 7	7	7	7	7	7	7	
PHYSICS	18	18	18	18	18	18	18
POL SCI	2	2	2	2	2	2	2

Outcomes
1. Fundamental Knowledge
2. Ill-Defined Problems
3. Communicate Effectively
4. Independent Learners
5. Work With Others
6. Intellectually Curious
7. Military Professionals

Attachment 2 (p. 2 of 3)

Number of Responses for Each Division

	OUTCOME						
	1	**2**	**3**	**4**	**5**	**6**	**7**
basic sci	87	86	86	86	86	85	86
engineering	56	56	56	56	56	55	55
humanities	43	43	43	43	43	43	43
social sci	31	31	31	31	31	31	31

Average Rating by Course for Each of Seven Outcomes

	OUTCOME						
	1	**2**	**3**	**4**	**5**	**6**	**7**
AEROENGR 215	2.50	1.25	1.50	1.50	1.75	1.50	1.75
ASTRO 320	2.80	2.20	1.80	2.60	1.60	2.20	1.80
ASTRO 410	2.86	1.14	1.86	1.57	1.71	2.00	2.57
BEH SCI 110	3.00	2.67	2.50	2.33	2.50	2.83	1.83
BEH SCI 310	2.31	2.56	2.75	1.88	2.75	2.13	2.75
BIOLOGY 215	3.00	1.17	1.83	1.67	2.50	1.80	1.83
CHEM 141	2.60	1.07	1.40	1.40	2.13	1.47	1.60
CHEM 142	2.63	1.26	1.32	1.47	2.11	1.47	1.58
CIV ENGR 310	2.00	2.80	2.55	1.65	3.00	1.50	2.20
COMP SCI 110	3.00	2.33	1.50	1.83	1.50	1.67	2.17
ECON 221	2.40	2.20	1.20	1.60	1.60	2.00	1.20
EL ENGR 215	2.60	1.00	1.60	1.60	2.00	1.00	1.00
EL ENGR 231	3.00	2.00	1.00	2.00	2.00	1.00	2.00
ENGLISH 111	2.64	2.09	3.00	1.91	1.82	2.27	1.91
ENGLISH 211	1.93	2.71	2.93	2.50	1.93	3.00	1.79
ENGLISH 311	2.00	2.20	3.00	2.20	2.00	2.20	2.80
ENGR 110	3.00	3.00	3.00	2.00	3.00	1.00	2.00
ENGR 310	2.80	1.00	1.00	1.20	1.80	1.40	1.20
ENGR 410	1.62	3.00	2.77	2.38	2.92	2.00	2.25
ENGRMECH 120	3.00	1.25	1.50	1.50	2.50	1.50	1.50
FRENCH 142	2.50	1.50	2.75	1.75	3.00	2.00	1.75
HISTORY 101	3.00	1.00	3.00	1.00	2.00	2.00	3.00
HISTORY 202	3.00	2.00	3.00	2.00	1.00	2.00	3.00
LAW 320	3.00	2.67	2.33	2.00	1.50	2.00	2.33
LAW 420	2.50	2.67	2.67	1.83	1.33	2.33	2.67
MATH 141	2.57	1.29	1.43	1.86	2.14	1.71	1.71
MATH 142	2.90	1.30	1.80	1.90	2.20	1.70	1.80
MATH 220	2.60	1.40	2.00	2.00	2.80	1.60	1.40
MGT 210	2.00	1.00	2.00	1.00	3.00	2.00	1.00
PHILOS 310	2.14	2.86	2.00	1.71	1.43	2.14	2.86
PHYSICS 110	2.70	2.40	1.60	2.10	2.50	2.10	1.80
PHYSICS 215	3.00	2.00	1.13	2.13	2.25	2.38	1.63
POL SCI 211	3.00	3.00	2.50	2.00	1.00	2.00	3.00

Attachment 4 (p. 1 of 1)

Average Rating by Department for Each of Seven Outcomes

	OUTCOME						
	1	**2**	**3**	**4**	**5**	**6**	**7**
AEROENG	2.50	1.25	1.50	1.50	1.75	1.50	1.75
ASTRO	2.83	1.58	1.83	2.00	1.67	2.08	2.25
BEH SCI	2.61	2.61	2.64	2.07	2.64	2.43	2.36
BIOLOGY	3.00	1.17	1.83	1.67	2.50	1.80	1.83
CHEM	2.62	1.18	1.35	1.44	2.12	1.47	1.59
CIV ENGR	2.00	2.80	2.55	1.65	3.00	1.50	2.20
COMP SCI	3.00	2.33	1.50	1.83	1.50	1.67	2.17
ECON	2.40	2.20	1.20	1.60	1.60	2.00	1.20
EL ENGR	2.71	1.29	1.43	1.71	2.00	1.00	1.29
ENGLISH	2.20	2.40	2.97	2.23	1.90	2.60	2.00
ENGR	2.00	2.47	2.32	2.05	2.63	1.78	1.94
ENGRMECH	3.00	1.25	1.50	1.50	2.50	1.50	1.50
FRENCH	2.50	1.50	2.75	1.75	3.00	2.00	1.75
HISTORY	3.00	1.50	3.00	1.50	1.50	2.00	3.00
LAW	2.67	2.67	2.56	1.89	1.39	2.22	2.56
MATH	2.73	1.32	1.73	1.91	2.32	1.68	1.68
MGT	2.00	1.00	2.00	1.00	3.00	2.00	1.00
PHILOS	2.14	2.86	2.00	1.71	1.43	2.14	2.86
PHYSICS	2.83	2.22	1.39	2.11	2.39	2.22	1.72
POL SCI	3.00	3.00	2.50	2.00	1.00	2.00	3.00

Outcomes
1. Fundamental Knowledge
2. Ill-Defined Problems
3. Communicate Effectively
4. Independent Learners
5. Work With Others
6. Intellectually Curious
7. Military Professionals

* NOTE: The display of this data is really by course prefix...the only impact is upon the Engr course, which really belong to "home" departments

Attachment 5 (p. 1 of 1)

Average Rating by Division for Each of Seven Outcomes

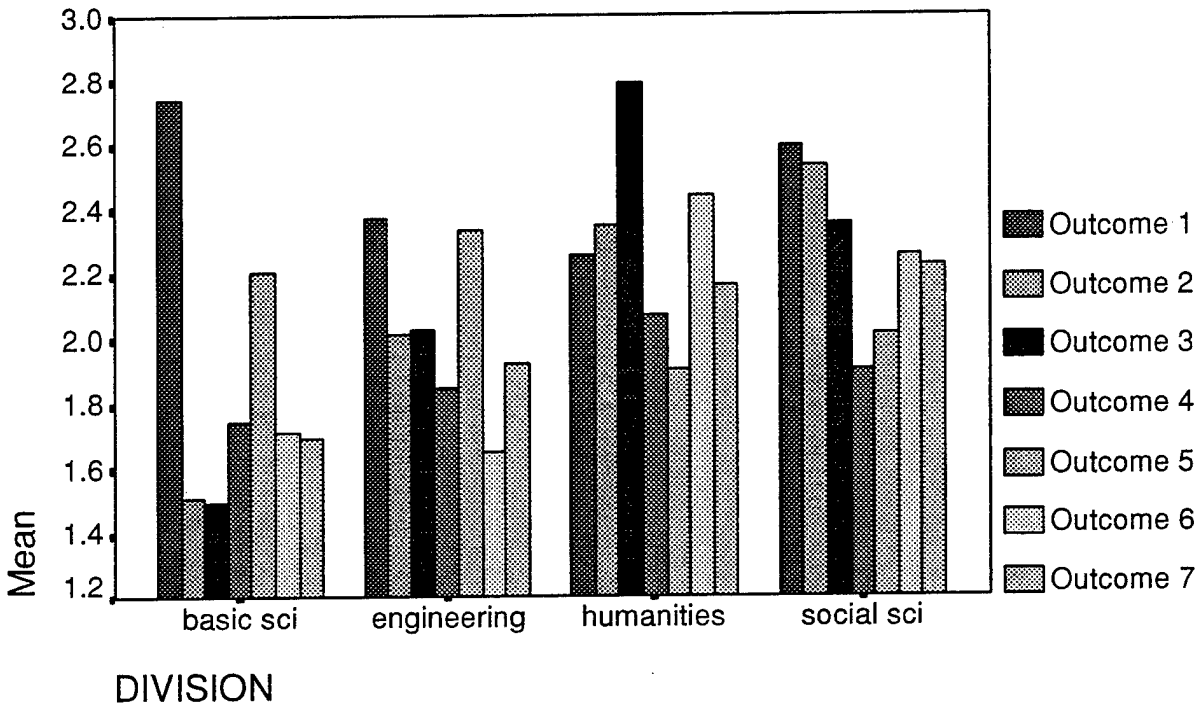
	OUTCOME						
	1	**2**	**3**	**4**	**5**	**6**	**7**
basic sci	2.75	1.51	1.50	1.74	2.21	1.72	1.70
engineering	2.38	2.02	2.03	1.85	2.34	1.65	1.93
humanities	2.26	2.35	2.79	2.07	1.91	2.44	2.16
social sci	2.60	2.53	2.35	1.90	2.02	2.26	2.23

Attachment 6 (p. 1 of 1)

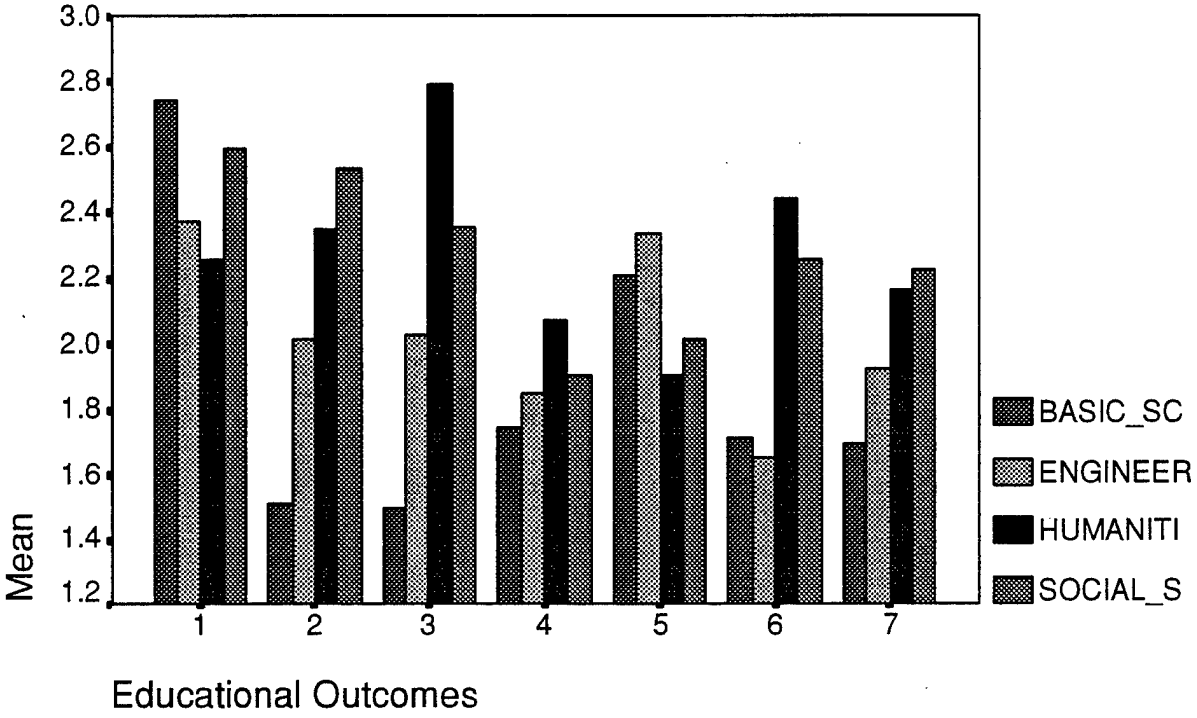
DF-Wide Overal Average

Variable	Mean	Std Dev	Min	Max	Valid N
CS4	1.86	.73	.0	3.0	216
CS6	1.93	.73	.0	3.0	214
CS7	1.93	.84	.0	3.0	215
CS2	1.96	.90	.0	3.0	216
CS3	2.02	.81	.0	3.0	216
CS5	2.16	.68	.0	3.0	216
CS1	2.53	.61	1.0	3.0	217

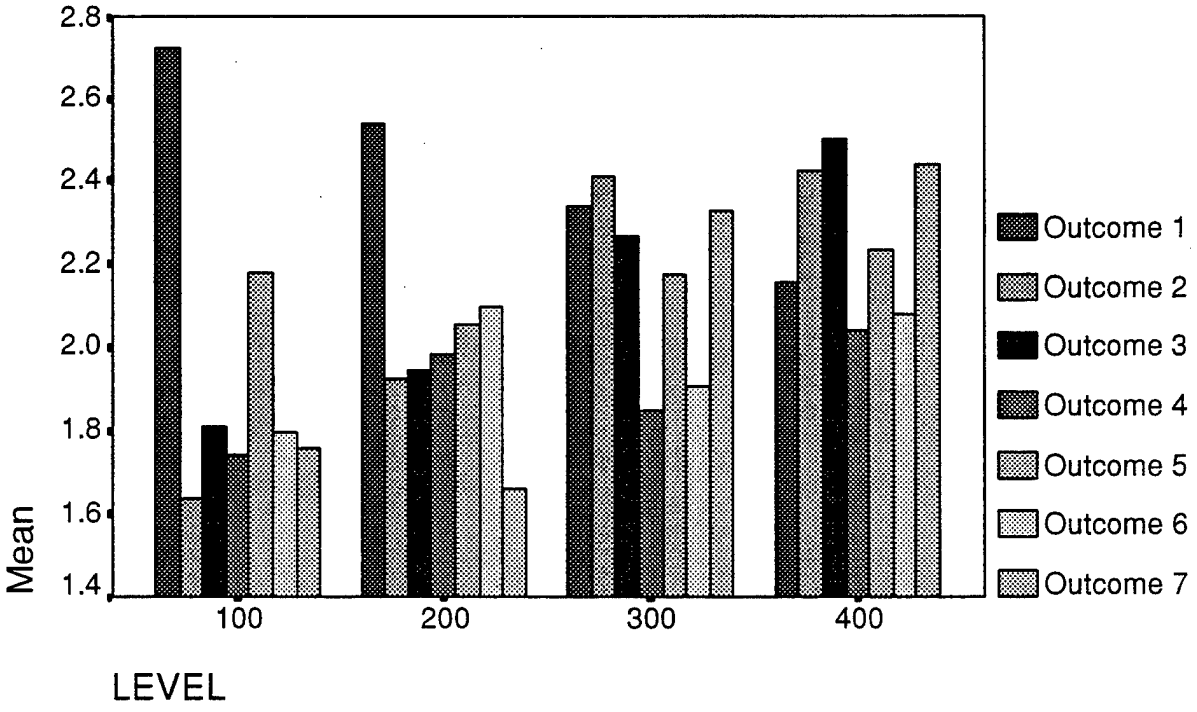
Attachment 7 (p. 1 of 2)

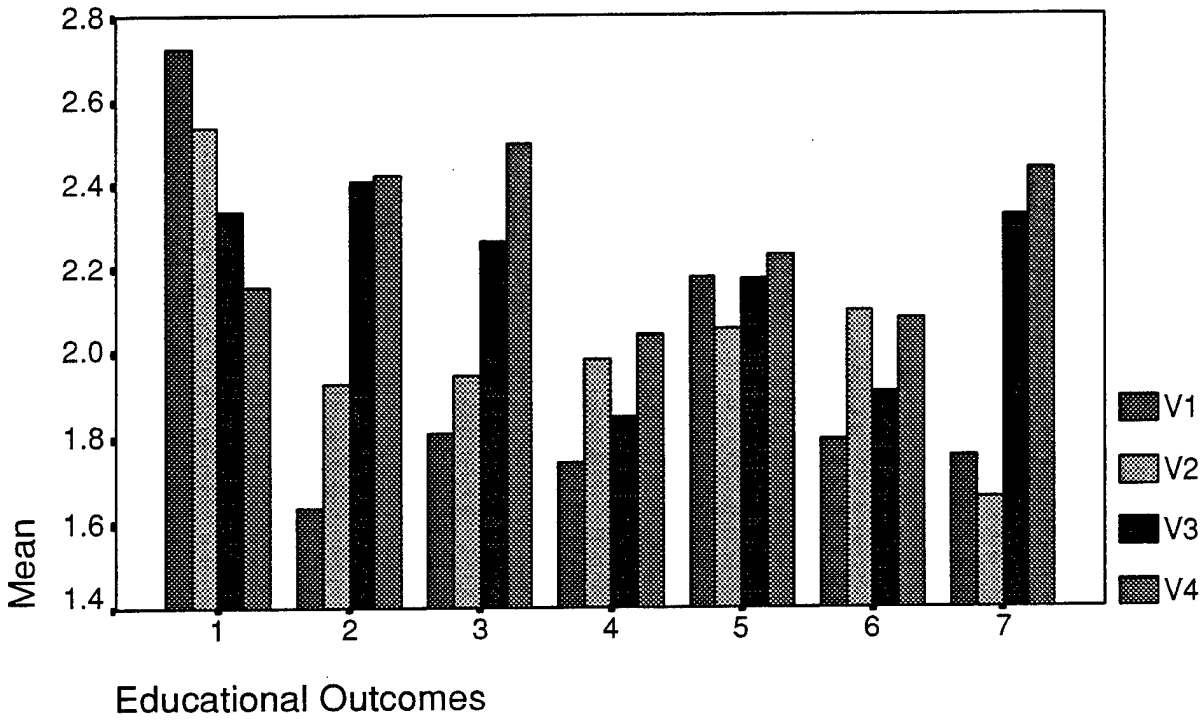


Attachment 7 (p. 2 of 2)



Attachment 8 (p. 1 of 2)





Outcomes
1. Fundamental Knowledge
2. Ill-Defined Problems
3. Communicate Effectively
4. Independent Learners
5. Work With Others
6. Intellectually Curious
7. Military Professionals

Student Critical Thinking Survey

Critical thinking skills do not develop in a vacuum. There are some classroom environments which actively promote the development of critical thinking abilities and others which inhibit it. In fact, this is the basic assumption underlying Moore's widely accepted Learning Environments Preference instrument used to yield cognitive development scores with respect to Perry's Model of cognitive development. Noreen and Peter Facione are two other notable pioneers in the area of thinking skill development and its assessment. Numerous conversations with the Faciones made clear the necessity of asking students about their classroom environments in order to assess the likely affects on their problem solving and critical thinking ability.

Based on ongoing research conducted by the Faciones at other universities, a 15 item questionnaire was developed to collect additional information about core course contributions to the educational outcomes. The first 11 items focused specifically on classroom climates conducive to the development of higher level thinking skills. The last 4 items represent a home-grown effort to address student perceptions of the integrated knowledge outcomes in terms which are within students' knowledge and experience. Although not specifically addressed, the assumption was made that cadets' intellectual curiosity (viz. attitudes) would be reflected in their overall responses. All responses employed a 5-point Likert Scale: 1-Strongly Disagree; 2-Disagree; 3-Neutral; 4-Agree; and 5-Strongly Agree.

Overall responses to the 11 questions generally fall between Neutral (3.0) and Agree (4.0). The highest average ratings deal with the instructor's role (The instructor encouraged thoughtful exploration of the central ideas and relationships in the course content (3.72), and the way the instructor conducted this course illustrated how to think in reasonable, objective, and fair minded ways (3.62)). In contrast the overall lowest rating (3.05) was garnered in response to the item: As a result of taking this course I am more fair minded. The next lowest score (3.29) reflected responses to the item: As a result of taking this course my interest and curiosity about the issues and questions in the subject area has grown. Many individual courses received overall average scores below neutral (3.0) in both these areas.

Response to the first three items concerning integrated knowledge are all very near Neutral (3.0). Cadets seem to neither agree nor disagree with statements about the relationship between the content covered in different core courses. Such connections continue to be a challenge for junior faculty who've had few opportunities for interdisciplinary education. Cadets responses to the final item dealing with cadets' belief that the core course is contributing to an integrated body of fundamental knowledge is more positive (3.45).

Educational Outcomes Assessment Working Group

Course Assessment

The Dean's Educational Outcomes Assessment Working Group has been asked to assess the contribution of the core courses to student learning. As a student in this core course, provide us with your responses to the following questions. Please use a #2 pencil to enter your responses on the digitek.

1. For question 1, use the letters below and code the department for the course in which you are completing this questionnaire.

Basic Sciences	**Engineering**	**Humanities**	**Social Sciences**
DFB = a	DFAN = f	DFENG = k	DFBL = o
DFC = b	DFAS = g	DFF = l	DFEG = p
DFCS = c	DFCE = h	DFPFA = m	DFL = q
DFP = d	DFEE = i	DFH = n	DFM = r
DFMS = e	DFEM = j		DFPS = s
MAS = t			

For the questions 2-4, use the following letters to code the course number.

1 = a	4 = d	7 = g	0 = j
2 = b	5 = e	8 = h	
3 = c	6 = f	9 = i	

2. First digit of the course number:
3. Second digit of the course number:
4. Third digit of the course number:

5. What is your status in this course?

   a. Freshman b. Sophomore c. Junior d. Senior e. Instructor f. Course Director

6. What academic division is your declared (or prospective) academic major in?

   a. Basic Sciences b. Engineering c. Humanities d. Social Sciences

(If you are an instructor or course director, please answer the following questions as you think students in this course are most likely to respond.)

Use the following scale for questions 7 - 21:

a	b	c	d	e
Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree

7. In this course I learned useful strategies for approaching complex questions in a variety of reasonable ways.

8. In this course I frequently found myself actively engaged in thinking about difficult questions for which we still need to find answers.

9. In this course I improved my ability to evaluate new information and analyze the central ideas of this subject area.

10. In this course I improved my ability to give sound reasons for my beliefs and opinions regarding issues in this subject area.

11. As a result of taking this course I find I am more fair-minded.

12. As a result of taking this course my interest and curiosity about the issues and questions in this subject area has grown.

13. As a result of taking this course my thinking is more focused and systematic, at least in this subject area.

14. The instructor encouraged thoughtful exploration of the central ideas and relationships in the course content.

15. The way the instructor conducted this course illustrated how to think in reasonable, objective, and fair-minded ways.

16. The assignments (tests, readings, projects, papers, classroom activities) in this course frequently engaged me in complex thinking.

17. This course lends itself to ill-defined problems (issues without a single approved solution).

18. The knowledge I have learned in this course has helped me better understand material in another core course.

19. This course routinely uses knowledge I have learned in another core course.

20. I have noticed an attempt in this core course to integrate material with other core courses.

21. This core course contributes to the creation of an integrated body of fundamental knowledge.

Case Summaries

COURSEID		Strategies	Actively Engaged	Evaluate & Analyze	Sound Reasons	Fair-Minded	Interest & Curiosity	Focused & Systematic	Thoughtful Exploration	Instructor Illustrated Thinking	Assignments	Course Lends Itself to Ill-Defined Problems	Knowledge Helped in Another Course	Uses Knowledge From Another Course	Attempt to Integrate	Integrated Knowledge
DFENG 311	N	275	275	275	275	275	275	275	275	275	275	275	275	275	275	275
	Mean	3.16	3.37	3.48	3.69	3.12	2.91	3.13	3.53	3.49	3.45	3.47	2.76	3.07	2.84	3.32
DFF 132	N	47	47	47	47	47	47	47	47	47	47	47	47	47	47	47
	Mean	3.30	3.19	3.77	3.36	3.43	4.30	3.70	3.81	3.64	3.28	2.87	3.06	2.72	2.89	3.64
DFP 142	N	62	62	62	62	62	62	62	62	62	62	62	62	62	62	62
	Mean	2.97	2.90	3.53	3.13	3.23	3.48	3.32	3.55	3.52	3.26	2.81	2.85	2.77	2.76	3.74
DFH 101	N	400	400	400	400	400	400	400	400	400	400	400	400	400	400	400
	Mean	3.17	3.28	3.50	3.50	3.16	3.42	3.20	3.60	3.46	3.28	3.25	2.96	2.87	2.87	3.54
DFH 202	N	331	331	331	331	331	331	331	331	331	331	331	331	331	331	331
	Mean	3.35	3.41	3.66	3.56	3.18	3.57	3.37	3.82	3.73	3.52	3.41	3.44	3.40	3.44	3.70
DFL 320	N	252	252	252	252	252	252	252	252	252	252	252	252	252	252	252
	Mean	3.65	3.69	3.83	3.87	3.50	3.78	3.70	3.77	3.72	3.52	3.76	3.04	2.85	2.89	3.91
DFL 420	N	146	146	146	146	146	146	146	146	146	146	146	146	146	146	146
	Mean	3.88	3.75	3.99	3.95	3.65	3.90	3.75	4.07	4.06	3.74	3.86	3.28	3.45	3.27	3.89
DFM 210	N	284	284	284	284	284	284	284	284	284	284	284	284	284	284	284
	Mean	3.37	3.23	3.54	3.35	3.19	3.53	3.35	3.80	3.74	3.37	3.43	3.15	2.95	3.12	3.54
DFMS 141	N	48	48	48	48	48	48	48	48	48	48	48	48	48	48	48
	Mean	3.83	3.40	3.96	3.50	3.08	3.27	3.75	4.17	4.17	4.04	3.02	3.23	3.33	3.38	3.60
DFMS 142	N	110	110	110	110	110	110	110	110	110	110	110	110	110	110	110
	Mean	3.55	3.47	3.56	3.42	3.07	3.10	3.53	3.47	3.51	3.56	3.13	3.64	3.34	3.55	3.56
DFMS 220	N	129	129	129	129	129	129	129	129	129	129	129	129	129	129	129
	Mean	3.20	2.84	3.52	2.90	2.53	2.51	3.28	3.68	3.60	3.36	2.59	2.36	2.24	2.40	3.00
DFP 110	N	469	469	469	469	469	469	469	469	469	469	469	469	469	469	469
	Mean	3.74	3.75	3.73	3.49	3.19	3.50	3.58	3.88	3.74	3.87	3.64	3.60	3.61	3.61	3.70
DFP 215	N	389	389	389	389	389	389	389	389	389	389	389	389	389	389	389
	Mean	3.53	3.40	3.57	3.20	2.90	3.26	3.47	3.65	3.54	3.58	3.05	3.31	3.44	3.28	3.43
DFPFA 310	N	266	266	266	266	266	266	266	266	266	266	266	266	266	266	266
	Mean	3.71	3.98	3.67	3.66	3.20	3.23	3.36	3.89	3.74	3.45	4.14	2.70	2.54	2.58	3.42
DFPS 211	N	175	175	175	175	175	175	175	175	175	175	175	175	175	175	175
	Mean	3.71	3.82	3.85	3.96	3.53	3.79	3.59	4.09	3.85	3.74	3.69	3.41	3.30	3.28	3.88
DFPS 212	N	466	466	466	466	466	466	466	466	466	466	466	466	466	466	466
	Mean	3.40	3.75	3.65	3.73	3.26	3.58	3.42	3.87	3.68	3.52	3.71	3.25	3.21	3.12	3.65
ENGR 410	N	355	355	355	355	355	355	355	355	355	355	355	355	355	355	355
	Mean	3.43	3.46	3.30	3.29	2.90	2.59	3.06	3.37	3.42	3.16	4.03	2.66	3.22	3.30	3.15
Total	N	9892	9892	9892	9892	9892	9892	9892	9892	9892	9892	9892	9892	9892	9892	9892
	Mean	3.42	3.41	3.59	3.40	3.05	3.29	3.41	3.72	3.62	3.51	3.38	2.99	3.05	3.00	3.45

Case Summaries

COURSEID		Strategies	Actively Engaged	Evaluate & Analyze	Sound Reasons	Fair-Minded	Interest & Curiosity	Focused & Systematic	Thoughtful Exploration	Instructor Illustrated Thinking	Assignments	Course Lends Itself to Ill-Defined Problems	Knowledge Helped in Another Course	Uses Knowledge From Another Course	Attempt to Integrate	Integrated Knowledge
DFAN 110	N	34	34	34	34	34	34	34	34	34	34	34	34	34	34	34
	Mean	4.00	4.29	3.97	3.71	3.24	3.56	3.32	4.06	3.65	4.24	4.35	3.91	3.88	3.91	3.94
DFAN 215	N	339	339	339	339	339	339	339	339	339	339	339	339	339	339	339
	Mean	3.33	3.03	3.56	3.12	2.71	3.48	3.50	3.49	3.39	3.45	2.76	2.76	3.18	2.97	3.37
DFAN 310	N	275	275	275	275	275	275	275	275	275	275	275	275	275	275	275
	Mean	3.32	3.08	3.53	3.10	2.87	3.19	3.48	3.72	3.72	3.49	2.77	2.85	2.96	2.93	3.25
DFAS 320	N	170	170	170	170	170	170	170	170	170	170	170	170	170	170	170
	Mean	3.36	3.16	3.48	3.18	2.72	3.61	3.38	3.61	3.49	3.53	3.25	2.61	3.02	2.72	3.22
DFAS 410	N	207	207	207	207	207	207	207	207	207	207	207	207	207	207	207
	Mean	3.26	3.23	3.51	3.29	2.84	3.48	3.36	3.80	3.74	3.38	2.94	2.72	3.18	2.88	3.37
DFB 215	N	406	406	406	406	406	406	406	406	406	406	406	406	406	406	406
	Mean	3.08	3.32	3.42	3.33	3.09	3.48	3.23	3.80	3.68	3.17	3.03	2.71	2.68	2.66	3.56
DFBL 110	N	375	375	375	375	375	375	375	375	375	375	375	375	375	375	375
	Mean	3.54	3.66	3.69	3.73	3.40	3.62	3.47	3.95	3.78	3.58	3.78	3.05	2.83	2.90	3.57
DFBL 310	N	152	152	152	152	152	152	152	152	152	152	152	152	152	152	152
	Mean	3.80	3.69	3.63	3.73	3.56	3.67	3.50	4.09	4.11	3.48	3.94	3.32	3.18	3.20	3.73
DFC 142	N	871	871	871	871	871	871	871	871	871	871	871	871	871	871	871
	Mean	3.46	3.40	3.63	3.19	2.89	3.13	3.42	3.61	3.53	3.60	3.10	3.09	3.16	3.09	3.43
DFCE 310	N	329	329	329	329	329	329	329	329	329	329	329	329	329	329	329
	Mean	2.88	2.93	3.01	2.92	2.35	2.40	2.83	3.29	3.05	2.76	3.62	2.05	2.26	2.27	2.59
DFCS 110	N	491	491	491	491	491	491	491	491	491	491	491	491	491	491	491
	Mean	3.63	3.52	3.71	3.12	3.00	3.34	3.66	3.74	3.73	3.74	3.65	2.77	2.64	2.77	3.33
DFEE 215	N	336	336	336	336	336	336	336	336	336	336	336	336	336	336	336
	Mean	3.19	2.75	3.46	2.93	2.69	3.15	3.45	3.54	3.49	3.29	2.46	2.87	3.33	3.11	3.11
DFEE 231	N	82	82	82	82	82	82	82	82	82	82	82	82	82	82	82
	Mean	3.85	3.18	3.87	3.20	2.82	3.24	3.82	3.88	3.82	4.01	3.05	3.51	3.74	3.51	3.62
DFEE 311	N	57	57	57	57	57	57	57	57	57	57	57	57	57	57	57
	Mean	3.67	3.61	3.60	3.44	2.95	3.42	3.58	3.84	3.86	3.81	3.39	3.09	3.53	3.32	3.51
DFEG 221	N	295	295	295	295	295	295	295	295	295	295	295	295	295	295	295
	Mean	3.51	3.28	3.62	3.42	2.94	3.30	3.47	3.79	3.66	3.56	3.38	2.87	2.76	2.73	3.40
DFEG 310	N	86	86	86	86	86	86	86	86	86	86	86	86	86	86	86
	Mean	3.55	3.79	3.69	3.55	3.33	3.62	3.66	3.98	3.90	3.64	3.67	3.40	3.44	3.44	3.69
DFEM 120	N	332	332	332	332	332	332	332	332	332	332	332	332	332	332	332
	Mean	3.73	3.43	3.76	3.34	2.96	3.17	3.66	3.56	3.55	3.82	3.23	3.38	3.52	3.42	3.55
DFEM 120Z	N	17	17	17	17	17	17	17	17	17	17	17	17	17	17	17
	Mean	3.76	3.59	3.53	3.24	2.82	2.94	3.59	3.94	3.88	3.88	3.41	2.94	3.41	2.88	3.47
DFENG 111	N	389	389	389	389	389	389	389	389	389	389	389	389	389	389	389
	Mean	3.38	3.46	3.72	3.82	3.30	3.12	3.48	3.81	3.71	3.66	3.45	3.14	2.94	2.93	3.55
DFENG 211	N	445	445	445	445	445	445	445	445	445	445	445	445	445	445	445
	Mean	3.18	3.44	3.54	3.60	3.05	2.95	3.19	3.75	3.44	3.56	3.63	2.62	2.65	2.54	3.27

Other Data Sources

The initial plan was to allow each of the assessment teams to cast their assessment nets broadly, seeking additional relevant information, and share the results of what they gathered with other teams. Divisional Assessment Teams were to be the primary contacts with departments and individual course directors. Departmental representatives could coordinate with individual course directors and faculty members to procure course syllabi, examples of student projects and products, and general comments concerning course philosophy, administration and plans for future changes. Similarly, Horizontal Teams were to have conducted additional focus sessions with students and junior faculty members to improve understanding of apparent patterns in critique and attitudinal data provided by students and junior faculty. The plan was for all such information to be collected and shared. This information, collected in separate course portfolios, was to have been provided to both Divisional and Horizontal Assessment Teams. However, the delay in getting the project started, lingering uncertainty as to individual participation and the limitation placed on the number of man-hours required by the individual departments prevented this plan from being executed. Some of these additional assessment activities did take place within teams but the information was not generally shared between the teams before the actual ratings were assigned. Similarly, the initial plan was to allow teams to meet and negotiate adjustments to discrepant scores. However, as the semester drew to a close, the high number of retirements, transfers and other duty obligations pending made it clear that simply getting the individual team reports completed was all that could be accomplished. Although the subsequent meta-analysis removed many of the discrepancies, the face-to-face negotiation and resolution would have been preferable if conditions would have permitted.

The general approach was to gather as much information as possible. Information that was available and potentially relevant was provided to the teams. Some of the information, such as the College BASE and Course Characteristics Survey, was reviewed and found by several teams not to be particularly helpful. As active participants in the existing educational system at the Academy, assessment team members also brought a rich store of personal experience, insight and anecdotes to the table. These less tangible resources are likely to have also strongly influenced the individual ratings teams assigned to the various courses. Encouraging the presentation and public consideration of these diverse views and opinions in the context of a great deal of quantitative data allowed teams to explore a variety of educational issues in considerable depth. Their willingness to engage one another with integrity and sensitivity is the foundation upon which this report is based.

Introduction to Part II - Divisional and Horizontal Assessment Reports with Commentary

Divisional assessment teams were provided with very similar information and also shared a common general understanding of education and the assessment process. Teams often included several individuals who had worked together on educational assessment for nearly 18 months. However, each divisional team was given nearly complete autonomy regarding how they conducted the assessment. The only requirement was for them to use any or all of the data provided to rate each core course's separate contributions to the three educational outcomes with respect to the rubric presented at the end of Chapter 1. Teams were encouraged to include any explanatory comments they deemed appropriate. Each report offers an account of real time assessment issues and dilemmas the teams faced and also illustrates alternative approaches to their resolution. Each team, as well as each member of the team, acted with complete integrity and commitment. However, the products did vary considerably in the extent to which they converged with the ratings of the same courses given by other teams. Additionally, unanticipated personnel changes left some teams (viz., social sciences and engineering divisional teams) with very little assessment experience or expertise.

The purpose of the commentary and critique at the end of each chapter is to draw out common themes and respond to some of the issues and concerns raised within the reports. The purpose is not an attempt to "second guess" or "correct" the work of any of the separate committees. Rather it is an effort to provide a general context for understanding and applying the work the committees did. It is also an opportunity to directly address some of the problems the teams encountered and suggest ways they might be avoided in the future. As the Educational Outcomes Assessment Working Group learned many times in their Phase I assessment activities: improvement is a much more useful goal than perfection, and the inability to achieve perfection is not a justification to quit trying or to quit assessing.

Commentary following each report will address three questions: 1) ***What has the report revealed about the three educational outcomes from the perspective of each academic division and level?*** 2) ***What has the report told us about the contributions of specific core courses in each division and level?*** 3) ***What can be learned from the various reports concerning our assessment process and practice in general?*** These questions will also be used as a framework for the meta-analyses in Chapters 10, 11 and 12.

Divisional Reports will be presented first, each with commentary immediately following the original report. Basic Sciences and Humanities Assessment Team reports will be presented first, followed by the reports of the Social Science and Engineering Divisional Teams. The three Horizontal Assessment Team Reports will then be presented and discussed starting with the 100-Level (freshman) courses followed the 200-Level (sophomore) and finally the upper level courses. In contrast to the divisional teams, an effort was made to exclude individuals with "vested interest" from horizontal teams (i.e., those who were from departments with core courses included in the courses to be assessed). Despite these distinctions, the high degree of convergence in the ratings given by the horizontal and divisional teams suggest that both used very similar criteria and attended to similar characteristics of the courses they observed.

Admittedly, horizontal teams had less disciplinary expertise than did divisional teams. This might be considered a weakness; however, such diversity also can add strength and generalizability to the overall assessment effort. Due to their extensive expertise in the subject matter, members of the divisional teams may have tended to focus more on what was taught rather than what was actually learned. Unfettered by such "subject matter expertise", horizontal teams might pay somewhat closer attention to what students actually learned or retained. One indication of this distinction is that all horizontal teams invited and included cadets as team members; divisional teams did not. It is also noteworthy that many of the junior faculty who participated as assessment team members were graduates of the Air Force Academy: although their masters' degrees may have been in philosophy or management, their undergraduate work had included many of the same technical core courses they were now evaluating.

Chapter 3

Basic Sciences Division Report

CHAPTER THREE - BASIC SCIENCES DIVISION REPORT

Dr. Donald M. Bird--Department of Chemistry
Lt Col Larry D. Strawser--Department of Chemistry
Maj Heather W. Scholan--Department of Biology
Lt Col William C. Hobart--Department of Computer Science
Maj Scott C. Dudley--Department of Physics
Maj James H. Rutledge--Department of Mathematical Sciences

The Basic Sciences Division Educational Outcomes Working Group representatives evaluated all nine divisional core courses based upon the data from student critiques, student survey, and the instructor survey. Each critique or survey was evaluated for student and instructor evaluation of each course's contributions to three educational outcomes: integrated fundamental knowledge, framing and resolving ill-defined problems and intellectual curiosity. The group considered both instructor and student evaluations and discussed each core course and its contributions before assigning a rating on the seven point scale.

The surveys clearly show that, in general, students rate the core courses from the Basic Sciences Division highest in integrating fundamental knowledge. As is pointed out several times in the following course discussions, the BSD instructors view this as their most important objective. Ratings are less than we would desire primarily because of the aspect of integration. While each department feels that integration is important, all of our core courses are at the 100 and 200 levels. Therefore, we feel very strongly that a major part of our mission is to help the students gain a body of fundamental knowledge that they will use and integrate into higher level courses. Just as it would be foolish to build a building on a foundation of sand, the members of our division feel it would be foolish to shift our focus to other outcomes at the expense of fundamental knowledge. As one of our members puts it, "How can we expect our students to show higher level critical thinking skills if they have nothing to think critically about?" Even though fundamental knowledge is our focus--and it will remain so--we still, based on the ratings, do a satisfactory job on the other two outcomes.

A couple of general comments are appropriate here regarding the Basic Science Division's view of framing and resolving ill-defined problems and of intellectual curiosity. Because of our focus on teaching foundational material, the instructors believe that we do not and should not place as much emphasis on these two outcomes. Our students did, however, consistently rate framing and resolving higher than the instructors--evidence that what is common knowledge to a practitioner of the discipline is very likely ill-defined to a novice. You will notice that intellectual curiosity has minimal comments from the represented core courses described in the following pages. Part of this is due to lack of emphasis in this division while much of the reason lies in lack of consensus on how one can measure such a subjective attribute. We strongly believe this is an outcome which should be measured for our students five to ten years after they leave the Academy. This could be done, for example, by surveying our graduates about the kinds of intellectual activities they pursued voluntarily after graduation.

This division would like to make one final comment about the data used to produce this report. We feel that much of the survey and critique data is very subjective, and as a result, difficult to draw direct, correlative conclusions about how we do our business and how we can improve. The departments have already embarked on their own assessments which they believe will produce data more meaningful for their individual missions. In the future, we should consider divisional surveys to gather appropriate assessment data rather than trying to lump all disciplines together and end up comparing apples to pomegranates.

Integrated Fundamental Knowledge

Comp Sci 110 - 4.5

This rating reflects the cadets' evaluation of the amount of material they learned in the course and their evaluation of how well integrated this material was with that in their other courses. The instructors for the course were unanimous in viewing this educational outcome as the most important of all seven outcomes and see a need to concentrate improvement efforts in this area. Based upon the cadet survey administered this semester, one area offering substantial opportunity for improvement is increasing the integration of other disciplines into the computer science curriculum.

However, improvements must be approached very carefully. The primary focus of CS 110 should remain the teaching of problem-solving techniques and the role of computers in enhancing our nation's warfighting capability. In the past, attempts to use problem domains from other disciplines in the programming portion of the course sometimes impeded the weaker students' understanding of programming concepts. Because they did not understand the problem domain in which the programming assignment was cast, the programming concepts appeared much more difficult than they actually were. The key to successfully meeting this challenge may well lie in structuring a programming exercise that reinforces important concepts in another discipline's core course and in return is well-supported by those teaching that course.

Math 220 - 4.5

Of the seven educational outcomes, Math 220 instructors considered this educational outcome to be the one that was stressed the most. Math 220 will be undergoing some minor changes Fall '96. The course will be called Math 300, reinforcing the idea that the course is for 2 degrees. Also, engineers will be encouraged to take Math 356, not Math 300. As a result, Math 300 will stress probability theory less and practical statistical applications more.

Biology 215 - 4.5

The Biology 215 instructors were unanimous in viewing this educational outcome as the most important of all seven outcomes. We all feel/felt that Biology 215 should strive for a 7. Due to the difference between the instructor (7) and cadet (4.6) views, we need to review this outcome. Based upon the cadet survey administered this semester, Biology 215 instructors need to evaluate using knowledge from another course and attempt to integrate this information into our course curriculum.

However, as with all core courses, this must be done very carefully. The primary focus of Biology 215 must remain the teaching of biology. This is the course where Biology majors lay

their foundation for courses to come and for those non-majors to appreciate and gain a better understanding for the biological world around them. The key to successfully meeting this challenge may well lie in working closely with other disciplines such as physics, chemistry, and statistics. Together, we should be able to incorporate material that would reinforce important concepts in all three discipline's core courses.

***Chemistry 141** - 5.0*

This rating is based primarily on the evaluation provided by the cadets enrolled in the course. This evaluation was based on students' perceptions of the amount of material they learned in the course and *their* evaluation of how well the material they learned in Chemistry 141 was integrated with their other courses.

It is important to note this outcome was viewed as the most important of the three outcomes under consideration by both instructors and cadets for this course. However, the instructors rated this outcome much higher than the cadets: it received an average value of 6.2. This fact, coupled with the cadet evaluation of 5.0, clearly dictates a need to focus our improvement efforts on the "Integrated Knowledge" outcome.

Knowing we need to focus our improvements efforts in this area, however, does not mean improvements can be easily accomplished. The biggest obstacle that instructors and cadets face in attempting to integrate fundamental knowledge learned in Chemistry 141 is the inescapable fact that cadets enrolled in Chemistry 141 have completed no other USAFA courses with which to integrate! In general, the only alternative instructors are left with is to "integrate" the information presented in class with some information that "will be learned" in the future. This frequently leaves cadets more frustrated than enlightened. Furthermore, attempts to "integrate" Chemistry 141 with other freshman courses proves frustrating for the instructor; no section contains students that are all enrolled in ANY other single course.

Nevertheless, we will make an effort to increase our baseline of 5.0 for this outcome. Obviously, we must carefully study and plan our improvement efforts and closely monitor their results to ensure that our efforts are working.

***Physics 110 and 215** - 5.0*

The instructors for this course viewed this educational outcome as the most important of all seven outcomes. The gap between instructors and students indicates that we have room for improvement in this area. One thing we are attempting to do to improve in this area is to establish "idea sections" with other departments, for example, a Math 152/Phys 110 section which has common students. Certainly, the students in this "idea" section should feel the two courses' material is more integrated, but the larger benefit comes from two instructors sharing their two disciplines, and then in turn sharing that with their respective departments.

***Math 141** - 5.0*

Of the seven educational outcomes, Math 141 instructors considered this educational outcome to be the one that was stressed the <u>most</u>. Math 141 will be undergoing dramatic changes Fall '96. The teaching philosophy is changing as well as the number of contact hours. It is believed that

these changes will better contribute to this educational outcome. A separate assessment plan for Math 141 has already been drafted, however, this rating may serve as an interesting benchmark.

Math 142 - 5.0

Of the seven educational outcomes, Math 142 instructors considered this educational outcome to be the one that was stressed the most. Math 142 will be undergoing a change in teaching philosophy this Fall. The course is adopting a drastically different course text that emphasizes concepts and applications and de-emphasizes computations and theory. It is believed that these changes will better contribute to this educational outcome. This rating may serve as an interesting benchmark.

Chemistry 142 - 5.5

This value is based primarily on the evaluation provided by the cadets enrolled in the course. Their evaluations were, in-turn, based on *their* perceptions of the amount of material they learned in the course and *their* evaluation of how well the material they learned in Chemistry 142 was integrated with their other courses.

Note that the cadet rating for this educational outcome for Chemistry 142 was significantly higher than for Chemistry 141. This is no doubt due to the fact that all cadets enrolled in Chemistry 142 had completed a semester at USAFA and had other course material to integrate with the information covered in Chemistry 142. In addition, cadets must *necessarily* integrate material learned in Chemistry 141 with Chemistry 142. Nevertheless, the cadet rating of 5.0 was still significantly lower than the instructor rating of more than 6.0. Thus, we clearly need to focus our improvement efforts on the "Integrated Knowledge" outcome for Chemistry 142 just as we do for Chemistry 141.

Improvements efforts for this educational outcome in Chemistry 142 should be much easier to accomplish than for Chemistry 141, however. There will be a variety of courses, in addition to Chemistry 141, that every cadet enrolled in Chemistry 142 has either taken or is taking with which material can be integrated. Obviously, we must carefully study and plan our improvement efforts and closely monitor their results to ensure our efforts are working.

Framing and Resolving Ill-Defined Problems

Math 220 - 4.0

This score was arrived at through a weighting of various questions administered to cadets. The questions focused on the ill-defined problems in the course. Of the seven educational outcomes, Math 220 instructors considered this educational outcome to be the one stressed the least.

Biology 215 - 4.5

The cadet perceptions of their problem-solving ability was 4.5. The instructors' goal for this educational outcome is 3.5. The instructors believe that problem-solving is not a predominant part of Biology 215. Biology 215 emphasizes a base of knowledge that requires more concept recognition than problem-solving. There are a few problem-solving opportunities given in our labs using the scientific method. Thus, the opportunity for improving this outcome lies in communicating to the instructors how the scientific method can be used to solve all problems.

Chemistry 141 - 4.5

As with the Integrated Knowledge outcome, this value is based primarily on cadet perceptions of their increased problem-solving ability. The average rating given to this educational outcome by the instructors teaching the course is only about 3.0, indicating that what seems "clearly" defined by the instructor is at least sometimes "ill-defined" by the cadet. That is, a well-defined problem for the expert can be an ill-defined problem for the neophyte. It appears that what it means to be an "ill-defined" problem is, itself, an ill-defined problem.

Attempting to show "improvement" in this outcome is problematical; should the improvement be a higher or a lower rating? Part of the difference in cadet and instructor perceptions with this outcome is clearly due to the fact that in the natural sciences and for a trained natural scientist working in his or her field--and certainly for the science introduced in a general chemistry course--there are no "ill-defined" problems! Thus, the ideal rating for Chemistry 141 would seem to be zero! Either we need to modify the definition of "ill-defined problem" to include "problem-solving ability" or we must increase our efforts to educate our students that chemistry is a "well-defined" natural science. In the former case, our objective rating then becomes 7.0; in the latter, our objective rating becomes zero.

An alternative would be to attempt to teach chemistry as if it were an "ill-defined" discipline: give the student only the information known in the days of alchemy, the eleventh century, for example, and ask him to "solve" his way towards the twentieth century. Given the fact that most USAFA cadets have had at least some exposure to fundamental science principles and given the constraints of time, however, such an approach seems ill-advised and inappropriate.

Chemistry 142 - 4.5

This value is the same as for Chemistry 141. As with the Integrated Knowledge outcome, this value is based primarily on cadet perceptions of their increased problem-solving ability. The average rating given to this educational outcome by the instructors teaching the course is analogous to Chemistry 141, about 3.5, clearly indicating that what seems "clearly" defined by the instructor is at least sometimes "ill-defined" by the cadet. That is, a well-defined problem for the expert can be an ill-defined problem for the neophyte. Furthermore, a previous course in chemistry did not mitigate this cadet perception, probably because Chemistry 142 presents a host of "new" ill-defined problems for the neophyte.

As with Chemistry 141, attempting to show "improvement" in this outcome is problematical: should the improvement be a higher or a lower rating? Using the same reasoning as was used for Chemistry 141, the ideal rating for Chemistry 142 should either be 7.0 and 0.0, depending on the interpretation given to this educational outcome.

Math 142 - 4.5

Of the seven educational outcomes, Math 142 instructors considered this educational outcome to be the one stressed least. As with Math 141, it is thought that the new teaching philosophy in the calculus curriculum will enhance this educational outcome.

Physics 215 - 4.5

Physics 110 - 5.0

Because the gap between instructor's views and student's view is not large, perhaps we are all right in this area. Nevertheless, we are attempting to improve in this area by creating real-world estimation problems. A major problem is lack of textbook support in this area. Good problems of this nature are simply hard to create. We are switching textbooks in the fall in part to bring in more of these sorts of problems.

Math 141 - 5.0

Of the seven educational outcomes, Math 141 instructors considered this educational outcome to be the one that was stressed the least. It is thought that the new teaching philosophy in the calculus curriculum will enhance this educational outcome. Part of the philosophy revolves around using real world problems to motivate the need for the calculus. Cadets will be allowed to struggle with a problem and develop their own ideas before a mathematical solution is offered. To the cadets, the problems are ill-defined. Most traditional calculus curriculums show the math and then the problem.

Again, a separate assessment plan for Math 141 has already been drafted, however this rating may serve as an interesting benchmark.

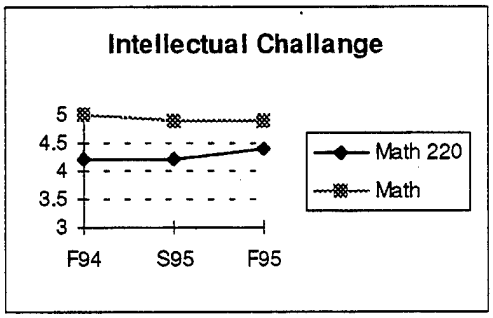
Comp Sci 110 - 5.0

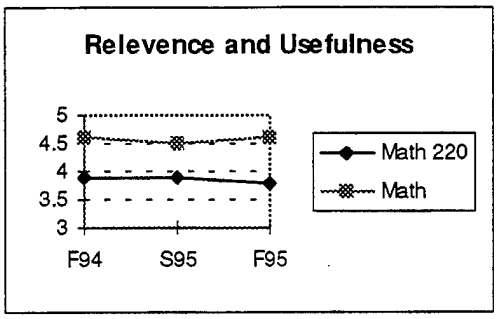
Again, this rating reflects cadet perceptions of their increased problem-solving ability. The instructors' overall rating indicates that an appropriate goal for this educational outcome is either 5.5 or 6. Problem-solving is already a predominant part of CS 110. Thus, the opportunity for improvement in this outcome lies in communicating to the cadets how the fundamental techniques taught in our core course can be applied in almost any domain. These techniques include framing the problem, devising a plan of attack, decomposing the problem into smaller subproblems, and then clearly specifying a solution. Another area for improvement may lie in increasing the cadets' confidence in their problem-solving ability. By better preparing cadets for their major programming exercise, we may be able to increase their success on this project, decrease the amount of help required from instructors to complete the project, and increase their overall feeling of accomplishment.

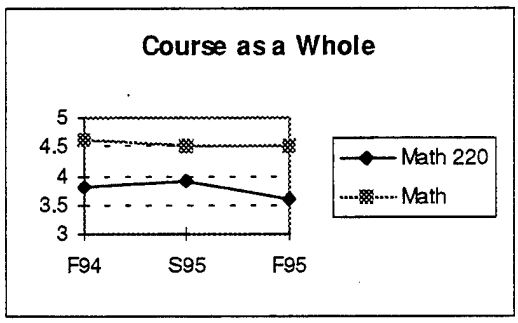
Intellectual Curiosity

Math 220 - 4.0

Math 220 cadet evaluations have recently (since Fall 94) been much lower than the department overall (see plots below). Instructors can offer several anecdotal reasons for this, however these reasons are not based on fact. Given the practicality and use of statistics in the Air Force, it seems that Math 220 should score better on the course critiques than they currently do. Perhaps some cadet focus groups could shed some light on why cadets don't rate Math 220 as high as they score other math courses. If the current restructuring of Math 220 does not result in an improvement, serious consideration should be given to conducting some experimental Math 220 courses to see if the course can be improved from the cadet perspective.







Comp Sci 110 - *4.5*

This rating is consistent with both the cadets' perception and the instructors' emphasis in the course. As such, we seem to be achieving our goals in this area, given the basic objectives of the course and its being offered in the fourthclass year.

Course changes being considered to increase the cadets' intellectual curiosity include increasing the amount of extra credit work available in the course to allow them to explore some topics more thoroughly, excusing top cadets from the final exam as an incentive to get them to do more than the minimum course requirements, and increasing the number of student presentations in the latter third of the course.

Math 141 - *4.5*

Of all the educational outcomes, this one is probably the most difficult to measure. Instructors did not stress this educational outcome very much in the course.

Math 142 - *4.5*

See Math 141 comments.

Biology 215 - 5.0

This rating is consistent with both the cadets' perception and the instructors' emphasis in the course. As such, we seem to be achieving our goals in this area, given the basic objectives of the course.

Physics 110 and 215 - 5.0

The gap between student and instructor's views is again not very large (zero for Phys 110). To stimulate student's interest we strive to bring in real-world examples of physics (the above mentioned textbook switch being part of this), as well as interesting actual demonstrations and labs. We have requested over $100k of equipment to be bought with potential fallout funds. We have received positive feedback from students in our recent improvements of labs and demos and we hope to continue this trend.

Chemistry 141 - 5.0

This rating is based mainly on the cadets' perception of the course. The instructors' rating was about 4.0. The relatively high cadet rating for this outcome is almost certainly due to the experience gained in the chemistry laboratory. In the Chemistry 141 laboratory, cadets can truly engage in "experiment." That is, innovation and creativity are encouraged after the fundamental objectives of a particular laboratory are met.

Unfortunately, not all students can participate in this creative or "intellectually curious" activity because they require the entire time given to perform the experiment to simply achieve the fundamental objectives. Moreover, cadets who **can** engage in creative activities frequently decline to do so, apparently because they do not find the rewards in doing so sufficiently motivating. Therefore, efforts to increase our cadets' intellectual curiosity would seem to lie either in making the entire experiment a "creative" activity (an endeavor we have ventured in the past with limited success) or in significantly increasing the incentives for a cadet to expand his or her "intellectual curiosity" (although such an activity seems at cross purposes: can a cadet be "forced" to become intellectually curious?).

We will have to carefully study which approach or combination of approaches will most likely improve this educational outcome and closely monitor our attempts to do so.

Chemistry 142 - 5.0

This rating is based mainly on the cadets' perception of the course, which was essentially the same as the cadet rating given to Chemistry 141. The instructors' rating also remained the same as the rating given to Chemistry 141, about 4.0. Again, the relatively high rating for this educational outcome is almost certainly due to the experience gained in the chemistry laboratory. Just as in Chemistry 141, in the Chemistry 142 laboratory, cadets can truly "experiment." That is, innovation and creativity are encouraged after the fundamental objectives of a particular laboratory are met.

Unfortunately, the same problems that were present in first semester chemistry exist in second semester: cadet capability to participate in creative or "intellectually curious" activity is not equally distributed and not all cadets who can engage in intellectually curious activity choose to do so.

Obviously, we will have to carefully study which approach or combination of approaches will most likely improve this educational outcome and closely monitor our attempts to do so. These attempts will be closely coupled with attempts to increase this educational outcome for Chemistry 141.

Basic Sciences Division Commentary & Critique

1) ***What has this report revealed about the three educational outcomes from the perspective of the Basic Sciences Division?***

Faculty in the Basic Sciences strongly endorse and closely identify with the *integrated knowledge* educational outcome. Six of the nine courses which make up the Basic Sciences core are taught in the freshman year and the other three are generally taught in the sophomore year. In nearly every course there was a clear faculty focus on providing the foundational knowledge necessary for further higher education. However, students didn't always appreciate this perspective or recognize the quantity and quality of knowledge the courses provided; it was also considered to be problematic to actually "integrate" current course material with information from future courses. This view suggests that later courses should receive higher integration ratings than those required to form the students' epistemological foundation.

In contrast to the close identification with the integrated knowledge educational outcome, there seemed to be considerable faculty resistance to the educational outcome dealing with students' ability to frame and resolve ill-defined problems. In every course (other than computer science), faculty self-reports suggested this outcome received the least emphasis. Ironically, however, students (in contrast to faculty) saw many of the problem solving techniques and approaches taught in Basic Science courses as being applicable to their ability to frame and resolve ill-defined problems. As this divisional report points out: perhaps courses that appear to be well defined to faculty experts are seen by students as being ill-defined. There may also be value in reacquainting faculty with the original framing and resolving rubric published in the Phase I: Initial Report (p. 15). For this rubric, the scientific method provided the initial scaffolding for the development of the nine characteristics which define excellence: (accurate problem identification; awareness of and compensation for personal bias; context awareness; application of external framing assumptions; systematic work; utilization of information from unsuccessful solution attempts; generating alternatives; reliance on known principles and skills; and displaying the appropriate level of confidence and commitment to solutions). In this case the students may be right, basic science courses just might contribute more to students' ability to frame and resolve ill-defined problems than their teachers realize. The assessment team is well aware of this conundrum and already actively engaged in educating their more recalcitrant colleagues.

Although apparently not as contentious and controversial as the ill-defined problem outcome, the intellectual curiosity outcome seemed to received only tepid endorsement from Basic Science faculty and staff. The suggestion from the Computer Science Department that increasing extrinsic rewards might be the best way to encourage intellectual curiosity is also somewhat problematic. There is fairly robust literature (as well as several local studies and research projects) that suggests external rewards and punishments are likely to diminish rather than encourage intellectual curiosity. If one assumes that intellectual curiosity is an attitude, lack

of emphasis on this outcome might well provide an explanation for some of the relatively low critique ratings from students of courses within this division. As shown in the Phase I Final Report and mentioned in Chapter One of this report, individual faculty member's relative emphasis on student attitudes is the single best predictor of students' ratings of their teacher's effectiveness and their own learning.

2) ***What has the report told us about the contributions of specific core courses in the Basic Sciences Division to the educational outcomes?***

The Basic Sciences assessment group rated the contributions of Basic Sciences core courses to be between 4.5 and 5.5 on a 7.0 point scale for integrated knowledge and to be between 4.0 and 5.0 for framing and resolving ability and intellectual curiosity. On the positive side, this suggests that the none of the courses were seen to be in real trouble. However, on the negative side, the lack of variance in ratings suggests that differences in these courses were not very salient. Nonetheless, 100 level Chemistry, Math and Physics courses received relatively high ratings across all three educational outcomes. Computer Science 110 was near the bottom in both knowledge and intellectual curiosity, but rated at the top of all Basic Science courses for its contribution to students' ability to frame and resolve ill-defined problems. If one accepts that the educational outcomes are at least somewhat positively related (as both past course critique and current assessment data suggest), such a disparity seems somewhat unlikely. However, given faculty resistance and confusion concerning the term "ill-defined" (and the team's hypothetical suggestion that a "0" rating might actually be a worthy goal for this outcome), assessment ratings in the ill-defined problem area may be less valid than those in other areas. Math 220, Statistics, received the lowest ratings in all areas. In discussing the course's low score for intellectual curiosity, the team suggests the use of cadet focus groups might provide further insight. An alternative approach might be to look at other (more positively rated) statistics courses taught as part of different academic majors. The case of Math 220 provides the first instance of a recurring theme in this report: when technical and non-technical versions of a technical core course are offered, the non-technical version will be rated by cadets (and assessed by teams) much more negatively than the high tech version. Perhaps this is self evident; it might also reflect the power of the self-fulfilling prophecy.

3) ***What can be learned from the Basic Sciences Division Assessment Report concerning the assessment process and practice in general?***

This team included several experienced educational outcomes assessors and the decision to examine the data as a group and build consensus is commendable. The overall average assessment ratings converged very closely with those provided by the two Horizontal Assessment Teams (HATs) that rated the same courses which is another good sign. It is also notable that none of the course ratings differed by 2.0 or more from the HAT ratings. However, the restricted rating range (1.0 for each outcome) reflects the difficulty the team had in discriminating between courses. This might be due to the fact that core courses typically involve at least 30 sections and nearly a dozen different faculty members; variation within the courses (from best to worst sections) may obscure overall differences between courses. This team also did not comment on other available objective data (such as the College BASE) which reflects both the relative strength of our students' knowledge of the Basic Sciences and improvement in many areas between the

freshman and senior years. Chapter 11 of this report contains an extended discussion of the results and potential relevance of the College BASE top several core courses and core course sequences. Although not definitive, such results do provide convergent support for the overall contribution patterns contained in these separate assessment reports. All in all, however, this report provides considerable evidence for the potential effectiveness of a division's efforts to conduct meaningful assessment of its own courses' contributions to these three educational outcomes.

Chapter 4

Humanities Division Report

CHAPTER FOUR - HUMANITIES DIVISION REPORT

Lt Col Mark D. Noe--Department of English
Lt Col Verner D. Mitchell--Department of English
Lt Col Charles W. Hudlin--Department of Philosophy and Fine Arts
Lt Col Judith E. Brisbois--Department of Foreign Languages
Major Tony T. Kern--Department of History

A. Brief discussion of Educational Outcomes as they apply to the academic division

The Humanities help officers understand what it is to be human in light of the continuity and diversity of human experience. In their basic forms, they teach the foundations of language and communication. They build on that basis by exploring the impact of events and ideas on our evolving conceptions of civilization--truly a task in framing and resolving ill-defined problems. Successful results require synthesizing knowledge from many disciplines and critically evaluating possible responses. The Humanities examine perennial ideas and immutable values which continue to describe and influence the human condition. They develop students' reasoning and communications abilities and encourage students to adopt a global perspective. They provide understanding of other cultures and languages. They explore critical concepts such as the nature of humanity, the nature of warfare, aesthetic appreciation in literature and the arts, the nature of change and evolution, the ethical life, and applications of these ideas to the military profession. They deal with these concepts in a manner which creates a sense of intellectual curiosity to inspire the continuous pursuit of knowledge.

1. Application to the Content (Integrated Knowledge) Outcome

a. A breadth of integrated, fundamental knowledge in the Humanities includes the following:

- A critical understanding of the worldviews, attitudes, values, and commitments of diverse cultures, especially our own; an awareness of the historical origins of various cultural traits
- The ability to read and understand complex, tightly reasoned English prose; the ability to speak and write clearly, precisely, and correctly in English
- The foundation and ability to understand and speak a foreign language well enough to satisfy basic survival needs in a foreign country
- The ability to think critically about human problems and goals; the ability to distinguish faulty from sound reasoning and to support one's beliefs with well-developed arguments

- An appreciation of the causes and consequences of war, especially as they relate to the human condition

b. Department-specific concerns include these:

- DFF: Learn about language through studying the target language; develop an understanding of culture through language
- DFH: Identify key geographic features and places of historic significance
- DFENG: Synthesize a wide range of source material--literary, cultural, historical, philosophical, political--to answer broad questions or argue an integrated thesis
- DFPFA: Become familiar with the achievements of Western philosophy, especially the rich intellectual history of thinking about morality in the West and the development of critical thinking

c. Individual course ratings:

- Foreign Language 131/2, 141/2, and 150: Ideal=**5**, Actual=**5**. The basic foreign language courses are designed to develop cadets' linguistic and cultural knowledge. Through study of the target language and culture, the cadets ameliorate their ability to understand and respect cultures that differ from their own by comparing and contrasting their own way of doing things with that of the target culture. Cadets are made aware of the target culture's general technological, social, political, and economic issues, without going into great detail.
- History 101: Ideal=**6**; Actual=**6**. History 101 is ideally suited to tie together course materials from a variety of disciplines through its analytical approach which uses the PIES model (political, ideological, economic, and sociological) to address changes across history. Cadets, instructors, and course directors agreed on the high rating in this area.
- History 202: Ideal=**5.5**; Actual=**5**. Military history integrates well with many of the Military Arts and Science courses offered by the CW side. In fact, many cadets voice concern over too much overlap. The focus of this course is on historical events and interpretation, rather than the art and science of applying the military tactics themselves. Instructors see this clearly, cadets have some trouble differentiating which results in some consternation (a.k.a., whining) about duplication of material.
- English 111: Ideal=**5**; Actual=**5**. Basic Composition is designed to help cadets write at a college level across the curriculum. But its approach is so basic that the students and faculty both see it as contributing only moderately to the entire curriculum. Obviously, the more successful the course is the more it affects other courses throughout a college career.
- English 211: Ideal=**5**; Actual=**4**. The Survey of World Literature teaches cadets to read and analyze texts and then report on those texts. It does so through the study of literary works from the last twenty-five centuries. Both instructors and students, however, view

the class in isolation, focusing on literature qua literature rather than on the universal applicability of the topic. Thus, though this number probably should be higher, it's only in the middle of the equation.

- English 311: Ideal=**6**; Actual=**5**. Advanced Composition and Public Speaking is meant to provide cadets with a final polishing of writing skills and a formalizing of basic speech skills, all achieved through the process of public argument on student-selected topics. While instructors try to present it as an argument course, basing success on written and spoken persuasive abilities, students often see only a repetition of other English (composition) and of Military Arts and Sciences (speech, though really only briefing) courses. They also find the grading troublesome--it's based on rhetorical skills, but they see it as purely subjective. They thus put little emphasis on the course, especially those who take it spring of senior year. While this course should prepare cadets for officer duties, they seldom see it doing so.

- Philosophy 310: Ideal=**6**; Actual=**4.5**. The course is not designed to impart knowledge. This is perhaps confirmed in the cadet perception, scoring 2.58 on attempts to integrate and 3.42 on contributes to integrated knowledge. It is designed to impart understanding and encourage reflection on difficult ethical problems largely created by human association and military operations. Consequently, the course is more likely to draw on knowledge cadets already possess as it becomes relevant to various moral issues. It is doubtful that cadets fully recognize this process or its value.

2. *Application to the Intellectual Curiosity Outcome*

a. Intellectual Curiosity in the Humanities includes the following:

- Officers who appreciate, apply to themselves, and continue to pursue throughout their lives the full range of literature, fine art, philosophy, and history
- DFF: Discover how language works, develop an understanding of cultural differences, and relate the two to current events
- DFH: Study the major events and players in world and American history
- DFENG: Individually research to locate persuasive and relevant materials which will decisively support an argumentative thesis on an open-ended topic; the most successful research of this sort requires intense interest in the subject matter and a curiosity to build on basic knowledge
- DFPFA: Explore, understand, and contribute to the on-going discussion of morality and ethics, especially as they apply to the profession of arms

b. Individual course ratings:

- Foreign Language 131/2, 141/2, and 150: Ideal=**5**, Actual=**5**.

- Cadets who are enrolled in Foreign Language 131 are at that level because they have either never taken a foreign language before, or have taken one for only one semester or less. Typically, these cadets have not already developed a great deal of intellectual curiosity toward language learning, or toward understanding cultural differences, and do not know how to relate these issues to their own experiences. Thus, at the 131 level, the relevance of language and culture studies is stressed, motivating the cadets to develop intellectual curiosity toward these studies.

- Cadets enrolled in Foreign Language 141 have already had two to four semesters of foreign language studies. These cadets are already curious about how language works, and about how other cultures operate. As in FL131, however, the relevance of linguistic/cultural knowledge is stressed in order to motivate the learners to further develop their intellectual curiosity toward these studies, and to see how they relate to an Air Force career.

- History 101: Ideal=**6**; Actual=**5**. The rapid pace of the course is frustrating to both instructor and cadet on occasion, but serves to act as a "tickler" for areas that stimulate cadet curiosity in a variety of areas, including social history, race and gender issues, military applications and tactics, and leadership. Unfortunately, since the course is taken during the fourth class year, follow up in the areas that are stimulated is difficult due to available time.

- History 202: Ideal=**6**; Actual=**5.5**. Cadets are fascinated by combat leaders, and continually seek greater detail on events and actions associated with those covered in class or readings. Instructors make optional and additional materials available, and provide adequate time for EI to answer these questions. Both instructors and cadets rate this area high.

- English 111: Ideal=**5**; Actual=**5**. Cadets must have some intellectual curiosity to succeed in Basic Composition, simply because the course requires some independent research, and instructors generally ask students to approach topics they (the students) find particularly interesting. Doolies, however, do not tend to research beyond a basic level--they don't have time to go beyond that level. Thus, though they might choose an item of interest, they have a limited opportunity to pursue that interest.

- English 211: Ideal=**6.5**; Actual=**6**. The Survey of World Literature is a bifurcated course when it comes to intellectual curiosity. Instructors rank it a solid 7, yet cadets put it at about 4 or 4.5. The instructors see literary study as positively necessary for future officers, as they must learn to understand and deal with personalities and their motives. Cadets apparently view the works of literature as depicting artificial human beings, however, and they thus denigrate the value of analyzing and explaining them.

- English 311: Ideal=**7**; Actual=**4.5**. An open-ended argument course, with a goal of persuading an audience on a topic of interest to and selected by the persuader, should generate a high level of intellectual curiosity. Cadets and faculty both rate this course rather low, however, and see it as drudgery (cadets) or technical skill (faculty) instead of as an opportunity.

- Philosophy 310: Ideal=6; Actual=4.5. Since the course focuses on moral issues and questions the cadets will face both as humans and as officers, it should do fairly well in this category. It does do relatively well in creating an intellectual challenge and encouraging independent thought (4.6). This places the course number one in the division and high among all departments, but there is room for improvement. The course scored 3.23 on interest and curiosity in subject area.

3. Application to the Framing and Resolving Ill-Defined Problems Outcome

a. Framing and Resolving Ill-Defined Problems in the Humanities includes the following:

- Officers who can apply the lessons of history to contemporary problems
- Officers who seek well-founded moral principles, apply them to daily life, and impose the resulting moral requirements on themselves
- Officers who appreciate the human dimension of war and behave accordingly
- DFF: Develop learning strategies; develop an understanding of target cultures through discussions of cultural similarities and differences and current events
- DFH: Understand the study of world history as an analytical process that is important to one's position as an American military professional in a global context; understand how historical events have produced today's world in terms of (a) the interaction of nations and people groups, especially the impacts of European and Western Culture on other world civilizations, (b) the impact of political, ideological, economic, and social developments on history, (c) the continuity and change in cultures, particularly in Europe, the Americas, Russia, Southwest Asia, Africa, and East Asia, and (d) the impact of individual personalities on historical events; and understand the importance of geography in world history
- DFENG: Read and respond to complex, open-ended, and ambivalent ways of representing the truth in the world; students must overcome their frustration that there is no single textbook answer to the questions these works raise
- DFPFA: Develop the skills necessary to understand, analyze, and evaluate competing moral positions on a wide variety of important topics; develop abilities to frame and resolve moral issues with analytical rigor; and apply course concepts to the moral problems officers confront in Air Force careers

b. Individual course ratings:

- Foreign Language 131/2, 141/2, and 150: Ideal=5, Actual=5. Even at the basic level, the cadets learn that there are many different ways to express the same thought, as well as many different yet valid attitudes toward life, work, education, etc. They learn communication fundamentals such as the reasons for and uses of different registers of

speech with various listeners, and the fact that words and expressions used in different contexts can reflect various shades of meaning. More importantly, they are exposed to the notion that other cultures may hold to different attitudes that are reflected in different ways of operating, and that these differences are perfectly valid.

- History 101: Ideal=**5**; Actual=**4.5**. As a foundational course in world history, this fast moving course does not lend itself well to delving deeply into problem resolving type discussions in the classroom, due to the vast amount of material that needs to be covered in 42 lessons. However, both cadets and instructors rated this course above average in this area. Linking common themes across a historical timeline appears to help identify common challenges and approaches for problem solving.

- History 202: Ideal=**6**; Actual=**5**. Combat decision making--a subject abounding with ill-defined problems--is studied in detail. The difficulty for instructors is to convince cadets--who often view the historical end as justification or repudiation of any decision--to see the challenges that faced the leader at the time of decision. Simulations and war gaming are common techniques to facilitate this dimension of cadet critical thinking.

- English 111: Ideal=**4.5**; Actual=**5.5**. This is a course in Basic Composition, and instructors tend to focus primarily on fundamental skills. Cadets, however, rate the open-ended research very highly, thus raising the course's actual rating for this category.

- English 211: Ideal=**6**; Actual=**6**. Decisions made by characters in novels, as well as the results of those decisions, really involve framing and resolving ill-defined problems.

- English 311: Ideal=**7**; Actual=**5.5**. Persuasive argument is founded on framing and resolving ill-defined problems, and this is a course in almost pure persuasion. However, instructors and cadets both too often view the process as more important than the result. In fact, the two are inseparable.

- Philosophy 310: Ideal=**7**; Actual=**5.5**. Of all the outcomes, framing and resolving ill-defined problems is the primary focus of ethics. Cadets are taught various problem-solving strategies for moral problems and are invited to think critically and creatively about the strategies and the solutions to various moral issues. The course is doing reasonably well on this outcome, but needs to constantly focus and improve in this area. It scored 4.14 on lending itself to ill-defined problems and a 5.16 on the course critique ill-defined problem outcome question.

B. Review of existing data relevant to the knowledge/content outcome

1: College BASE Data

- In the fall of 1990, freshman and senior cadets were tested with the standardized College Basic Academic Subjects Examination (BASE). "Subject scores . . . are reported on a scale ranging from 40 to 560, with an average being 300. Skill scores are reported as

High, Medium, or Low." No USAFA scores fell below the 90th percentile; most were 97 or higher. Here is a listing of the Humanities-related areas:

- English, overall: Freshmen/seniors averaged 323/346
  - Reading & Literature: Freshmen/seniors averaged 326/338
  - Reading Critically: Freshmen/seniors scored 26%/45% High, 71%/52% Medium, 3%/2% Low
  - Reading Analytically: Freshmen/seniors scored 34%/45% High, 55%/52% Medium, 11%/2% Low
  - Understanding Literature: Freshmen/seniors scored 21%/40% High, 76%/60% Medium, 3%/0% Low
  - Writing: Freshmen/seniors averaged 306/333
  - Writing as a Process: Freshmen/seniors scored 32%/55% High, 61%/42% Medium, 8%/2% Low
  - Conventions of Written English: Freshmen/seniors scored 21%/20% High, 74%/80% Medium, 5%/0% Low
  - Writing Exercise: Freshmen/seniors scored 0%/0% High, 87%/95% Medium, 13%/5% Low
- Social Studies, overall: Freshmen/seniors averaged 341/367
  - History: Freshmen/seniors averaged 336/353
  - Significance of World Events: Freshmen/seniors scored 37%/52% High, 63%/47% Medium, 0%/0% Low
  - Significance of U. S. Events: Freshmen/seniors scored 11%/40% High, 87%/60% Medium, 3%/0% Low
- Three general Competency scores, showing cognitive processing skills, regardless of content:
  - Interpretive Reasoning: Freshmen/seniors scored 97%/97% High, 3%/2% Medium, 0%/0% Low
  - Strategic Reasoning: Freshmen/seniors scored 66%/97% High, 34%/2% Medium, 0%/0% Low
  - Adaptive Reasoning: Freshmen/seniors scored 11%/22% High, 89%/75% Medium, 0%/2% Low
- Composite scores for this exam: Freshmen/seniors averaged 355/380
- Cadets showed increased ability in nearly every area, as one would expect after three years of education and maturation. Only two areas actually showed a drop in ability:

- Cadets fell off slightly in dealing with Conventions of Written English (21% to 20% at the high end), though there were no seniors placing at the low end.
- Cadets' Adaptive Reasoning ability increased at the high end (11% to 22%) but actually fell at the low end (0% to 2%).
- These drops are minuscule in comparison to the increases in all other areas. However, that these two particular areas do not show the strong increases of other areas may suggest a tendency among a few cadets to be less accurate writers and more lock-step thinkers. Even here, though, the numbers are so small as to be insignificant.

2. Historical Data from Engineering 410 Comprehensive Exam: Seniors in the Spring 1995 semester Engineering 410 course, a core course in Engineering Systems Design, took a test centered on an Air Force logistical problem. Results of this test do not reflect an objective measure of student ability to frame and resolve ill-defined problems with respect to an absolute standard, but they do suggest a general ability to deal with such problems. Fully 80% of the tested students clearly showed an ability to frame and resolve ill-defined problems, and this result was based on a test for which no student was specifically prepared. Results are not broken down by division.

3. Other Data: DFPFA conducted a similar comprehensive exam in their Philosophy 310 class. Review of the materials is still in process, but initial data suggests results similar to those found with the Engineering 410 exam. This study will serve only as a baseline for later work, as the department is also in the process of a complete revision of the course itself. This may be a useful tool to compare with a similar exam done once the new course is established.

C. Review of existing data relevant to critical thinking outcomes

1. Inventory of curricular practices

a. By division: See *Phase I: Final Report.*

- Framing and Resolving Ill-Defined Problems Outcome: Humanities courses often teach some basic factual material, but they generally get beyond that level rather quickly. The faculty designs courses to emphasize dealing with gray areas; students must synthesize information to reach defensible answers (rather than merely memorize and parrot back an approved solution).
- Intellectual Curiosity Outcome: Across divisions, faculty say essentially the same thing, expressing concern that cadets have a tendency toward intellectual curiosity, but that tendency is stifled due to time constraints--cadets appear to gain an interest in certain topics, but they have limited time to pursue the topics. Faculty members expect students to become interested enough in a topic that they explore beyond the assignment's formal requirements, research outside the textbook, and think beyond the narrow borders of the written task. They expect students to ask questions and not accept everything they hear

and read, to take things into their own hands and learn what makes them work. But cadets seldom do any of these things. Teachers assume this is because cadets don't have time.

b. **By course:** See *Phase I: Final Report.*

- Framing and Resolving Ill-Defined Problems Outcome: Faculty members believe cadets do not improve their abilities to frame and resolve ill-defined problems to the extent desired. Instructors who have tried to improve cadets' abilities in this area have often met resistance and resentment from students. Benefits from their efforts can be negated by such attitudes alone. The effectiveness of the faculty to teach cadets to frame and resolve ill-defined problems may increase as such efforts become more commonplace and hence more acceptable. Clearly, the faculty also needs to ensure upper-division courses are integrating work with ill-defined problems wherever appropriate.
- Intellectual Curiosity Outcome: Only one additional point needs be made about course versus division assessments of intellectual curiosity. Instructors say that 400-level majors courses have less inherent design features for stimulating intellectual curiosity than do 400-level core courses. The majors courses also do less to grade that effect. These results are opposite what the committee expected, though they do offer positive analysis for this survey--which is, after all, geared toward core courses.

2. Current critique results and recent trends:

- Faculty data: These data come from Attachment 3 to the 16 May 1996 Instructor Assessment of Core Courses. A score of more than 2.5 on the 3-point scale suggests a heavy emphasis on an individual outcome, as viewed by department faculty; a score of over 2 but no more than 2.5 suggests an above-average emphasis. Note that points are based on participation by 30 English faculty (with the fewest, 5, in English 311), 4 Foreign Language faculty (all from the same basic French course), 2 History faculty (one from each core course), and 7 Philosophy faculty.
- Outcome 1 (Fundamental Knowledge): History 101 and History 202, English 111, Foreign Language 142 all place heavy emphasis on fundamental knowledge; Philosophy 310 places an above average emphasis on this outcome.
- Outcome 2 (Frame and Resolve Ill-Defined Problems): Philosophy 310 and English 211 heavily stress framing and resolving ill-defined problems, while English 311 and English 111 are above average in their emphasis of this outcome.
- Outcome 3 (Communication): Nearly all Humanities courses emphasize the communication skills outcome; only Philosophy 310 ranks this lower, and even it scores the outcome at the 2 level.
- Outcome 4 (Independent Learners): English 211 and 311 stress independent learning rather heavily.

- Outcome 5 (Teamwork): Foreign Language 142 expects students to work together as a basis for success.
- Outcome 6 (Intellectual Curiosity): English 211 stresses intellectual curiosity heavily, while English 311 and 111 and Philosophy 310 place an above-average emphasis on it.
- Outcome 7 (Officership): History 101 and 202, Philosophy 310, and English 311 all focus on application of the course's work to the military profession.
- Thus, according to division faculty, Humanities Division core courses are primarily oriented toward the Communication, Intellectual Curiosity, and Framing and Resolving outcomes, in descending order of emphasis.
- Student data: These data come from the Course Assessment done in all core courses in April 1996. Students were asked 15 questions pertaining to three DF outcomes (Integrated Knowledge, Framing and Resolving Ill-Defined Problems, and Intellectual Curiosity). Scores of more than 3.5 indicate heavy stress on a particular outcome; scores of more than 3 to 3.5 indicate above-average stress on that outcome, and of more than 2.5 to 3 indicate slightly more than average stress on that outcome. Participation ranged from 275 to 445 students in English courses, 47 to 62 students in Foreign Language courses, 331 to 400 students in History courses, and 266 students in the Philosophy course.
- Integrating Knowledge: Students generally ranked 3 of the 4 questions in this category at slightly above the average, yet they consistently ranked the 4th (and overall) question fairly high. That is, students generally ranked lower (from 2.54 to 3.14) the use of material from this course in another course (18), the use of other course material in this course (19), and any visible attempt to integrate course material in this course with another course (20). In spite of these rankings, students ranked these courses higher (3.27 to 3.74) as contributors to "an integrated body of fundamental knowledge" (21). This generalization holds true for EVERY Humanities core course. This may suggest student misunderstanding of the questions; it may also suggest simple inconsistency on the part of the students.
- Framing and Resolving Ill-Defined Problems: Students ranked questions 7-11 and 13-17 (all concerned with variations in approach to dealing with knowledge lacking absolute answers) in these ranges:
    - English, 3.05 to 3.82
    - Foreign Language, 2.81 to 3.81
    - History, 3.16 to 3.82
    - Philosophy, 3.20 to 4.14

These ranges are reasonable, as the basic language courses start at the lower levels (2.81, or slightly above average stress on framing and resolving ill-defined problems), yet all courses go up to just over 3.8 (above average or heavy stress on framing and resolving ill-defined problems), and Philosophy climbs to well over 4 (heavy emphasis on framing and resolving ill-defined problems).

- Intellectual Curiosity: Question 12, asking about subject area curiosity, included responses in these ranges:
  - English, 2.91 to 3.12
  - Foreign Language, 3.48 to 4.30
  - History, 3.42 to 3.57
  - Philosophy, 3.23

3. *Faculty Survey results and Critique study:* *See* Phase I: Final Report.

- The Faculty Survey showed that courses using an approach which includes some emphasis on collaborative, two-way critical inquiry, especially at the upper-division level, tend to be more effective than courses employing a traditional didactic approach. Upper-level and non-technical courses also tend to encourage the development of reflective thinking. And such courses tend to score better in course critique ratings. By extension, departments with more teachers who engage in this kind of teaching activity tend to have higher department course critique averages. Higher-level Humanities division courses, then, should score well on course critiques.

4. *Other data: None.*

D. Review of existing data relevant to other outcomes:

1. *Current critique data and trends:*

- The spring 1996 survey of core course faculty confirmed some general beliefs about Humanities Division core courses. Generalizing about 100 and 200 level courses, they accent three outcomes: fundamental knowledge, communication, and intellectual curiosity. The higher-level courses focus on framing and resolving ill-defined problems, intellectual curiosity, and officership outcomes. Using division-wide averages, the primary outcome focused on by Humanities instructors is communication. It's followed by intellectual curiosity and framing and resolving ill-defined problems. These show a substantial difference from DF-wide averages, which rank fundamental knowledge and working effectively with others as the primary faculty concerns. At least one divisional role shows through in these results--Humanities strengthens a different side of cadets, pushing them to think and function in ways other core courses don't. Though a generalization, it clearly represents the view of divisional faculty.

2. *Other data:*

- Critical thinking questions: None.

- New course characteristics: None.
- New faculty priorities for teaching: None.

Humanities Division Commentary & Critique

1) ***What has this report revealed about the three educational outcomes from the perspective of the Humanities Division?***

The Humanities Divisional Assessment Team viewed the three educational outcomes as being closely related to each other and recognized that strong contributions to any one outcome were contingent on solid contributions to the other two. For example, the acquisition of integrated knowledge actually involved thinking and integrative skills and the skill of framing and resolving ill-defined problems required the patience and persistence provided by the development of intellectual curiosity.

Two significant obstacles to cadets' development of these outcomes were noted. One involved cadets' own lack of experience and knowledge: the team asserted that both topics and skills presented in literature and philosophy courses are broadly applicable to life in general, however, students' lack of experience and wisdom prevent them from realizing these connections. In the team's view, another significant impediment to cadets' internalization of the desired educational outcomes is ever-present time pressure. Emphasis on quick thinking and timeliness leave little time for authentic reflection or complex analysis. We often require students to submit products that look like logical arguments without giving them the time and feedback needed to really develop the underlying skills. The most negative effects of our ubiquitous emphasis on speed (instead of accuracy, quality or depth) is on our students' intellectual curiosity. More than any other division, the Humanities faculty have tried to offset the negative effects of excessive time pressure by providing students many choices regarding class assignments.

2) ***What has the report told us about the contributions of specific core courses in the Humanities Division to the educational outcomes?***

The Humanities Division Working Group chose to first establish an *Ideal* rating for each course and then use the data available to derive an *Actual* assessment score. Such a process seems a logical extension of the recognition that every course cannot (and perhaps should not) contribute equally to each and every educational outcome. However, this procedure does create some difficulties in identifying which particular courses offer the greater opportunities for improvement. The identification of *Ideals* which reflect a less than 7 rating also seem to create confusion about what the ratings really mean. The rating descriptors initially agreed upon by the overall Phase II Assessment team included the following descriptions following ratings of 5, 6, and 7:

7 - Best possible contribution; reflects state-of-the-art educational philosophy and practice; consistently solid implementation across all sections of the course.

6 - Excellent contribution; no deficiencies noted; solid consistency across sections; substantial evidence of appropriate policies and sound teaching practice.

5 - Slightly Above Average contribution; certain policies or practices may be particularly noteworthy but consistency across sections or course segments could improve.

It may be difficult for some faculty members to understand what message is being conveyed by listing an *Ideal* rating of "5" instead of 7. The inclusion of relative terms such as "*Best possible*" was intended to allow assessors the latitude they needed to evaluate even those outcomes that were not considered to be a course's top priority. Identifying separate non-maximal ideals also seems to run contrary to the interconnected nature of the educational outcomes elaborated by this assessment team earlier.

Nonetheless, the assessment scores provided suggest individual strengths and weaknesses of particular courses. History 101 received the top mark (6.0) for its contributions to cadets' integrated knowledge. English 211 was selected as the top contributor to both the ability to frame and resolve ill-defined problems and development of intellectual curiosity. English 111 had the unique distinction of being seen as contributing more to cadets' intellectual curiosity (5.5) than was considered ideal (4.5).

At the other end of the continuum, English 311 was assessed to be a total of 5.0 points below its ideal and Philosophy 310 garnered a cumulatively deficient 4.5 rating points across the three educational outcomes. In part, this reflects extremely high ideal ratings (7s in intellectual curiosity for both courses and a 7 in ill-defined problems for English 311). However, the 4.5 ratings both courses received in absolute terms were also the lowest in the division. The 4.0 rating of contribution to integrated knowledge awarded to English 211 and the 4.5 rating for contribution to cadets' ability to frame and resolve ill-defined problems might also represent significant opportunities for improvement.

3) ***What can be learned from the Humanities Division Assessment Report concerning the assessment process and practice in general?***

This assessment report reflects a very thorough, systematic and methodical approach. The problem of restricted ranges noted in the Basic Sciences Divisional Assessment appears not to have been a problem in the Humanities. The range of *Actual* scores was 2.0 for *integrated knowledge* and 1.5 for both *ill-defined problems* and *intellectual curiosity*. Divisional ratings seemed to be fairly consistent with those provided by the 3 Horizontal Assessment Teams. In fact, of the 24 (8 courses X 3 outcomes) possible comparisons between the two ratings, only one differed by 2 full points. This disagreement occurred in English 211 (the divisional team rated the contribution to intellectual curiosity as 6.0 but the horizontal team rated this contribution as only a 4.0).

Although the ratings discriminated among the courses' contributions and converged with those provided by the horizontal assessment team, the explanations often did not contain specific evidence to justify the ratings. Although student and faculty ratings appear to have received considerable attention, other data such as the College BASE and curricular characteristics seem to have been largely ignored. As with the Basic Sciences, College BASE results suggest as

freshmen, cadets score relatively well in the Humanities and improve significantly in nearly every area. The Humanities' faculty's focus on attitudes and intellectual curiosity is reflected in providing cadets with more curricular choices than in other divisions and this appears to have contributed to relatively high student critique ratings for the division.

Chapter 5

Social Sciences Division Report

CHAPTER FIVE - SOCIAL SCIENCES DIVISION REPORT

Lt. Col Justin D. Rueb—Department of Behavioral Sciences
Major Richard O. Abderhalden—Department of Management
Major Gwendolyn M. Hall—Department of Political Science
Major Timothy W. Murphy—Department or Law
Captain Andrew P. Armacost—Department of Management
Captain Ann L. Gorney—Department of Economics and Geography

INTRODUCTION

This report draws conclusions for the Social Sciences Division based on several surveys conducted by the Dean's Educational Outcomes Assessment Working Group, the Horizontal Assessment Teams, and the Division Assessment Working Groups during Spring Semester 1996. Although these surveys provide information covering all divisions and all core courses, this report is concerned only with the Social Sciences Division.

To facilitate this discussion the following outline is used. After this introduction:

- The four principal surveys are discussed with a brief comparison of the Social Sciences Division to the rest of the Academy.
- The individual Social Sciences Division courses are rated with respect to the rest of the Social Sciences Division.
- General conclusions about the Social Sciences Division are drawn and discussed.

The initial Phase II Charter for the divisional working group was amended. Initially our charter was "...to review all existing assessment data and conduct analyses to assess the contributions of each of the core courses offered within their respective academic divisions to each of the educational outcomes." It became apparent early in the semester there were large differences within and across courses and divisions as to what, if any, data was available. With this in mind, the current semester became as much a data gathering semester as a data analysis semester. To better balance this work only three of the Dean's seven Educational Outcomes were examined. These were outcomes one, two, and six as listed below.

1. Officers who possess breadth of integrated, fundamental knowledge in the basic sciences, engineering, the humanities, and social sciences, and depth of knowledge in an area of concentration of their choice.
2. Officers who can frame and resolve ill-defined problems.
6. Officers who are intellectually curious.

SPRING 1996 SURVEYS

The following four surveys were conducted during Spring Semester 1996. Each is designed to study the Dean's Educational Outcomes and to determine whether the core courses are accomplishing them.

Course Assessment Survey

The course assessment survey allowed course directors, instructors, and students to rate their course's contribution to DF educational outcomes 1, 2, and 6.

Methodology: Questions 7-16 are from the Facione Survey and will be used as a comparative measure in a separate report. The most direct measures of the Dean's Educational outcomes are Question 12 as a measure of intellectual curiosity (DF Educational Outcome 6), Question 17 as a measure of framing and resolving ill-defined problems (DF Educational Outcome 2), and questions 18-21 are used to rate integrated knowledge (DF Educational Outcome 1). Items 18-21 were averaged to produce a single number for comparison. These measures for the three educational outcomes were collected for all core courses at USAFA. The average of the Social Sciences Division's courses were compared to the DF average.

Assessment of the Social Sciences Division: The Social Sciences Division was higher than the DF average for all three of the educational outcomes. Fundamental knowledge was just slightly above the DF average and both framing and resolving ill-defined problems and intellectual curiosity were significantly higher as shown below.

Educational Outcome	DF Average	SS Average
1. Fundamental Knowledge	3.25	3.28
2. Framing and Resolving Ill-defined Problems	3.40	3.70
3. Intellectual Curiosity	3.35	3.64

DF End of Semester Critique Questions With Respect to DF Educational Outcomes

The DF end of semester critique completed by all students was used to rate each core course's contribution to DF educational outcomes 1, 2, and 6.

Methodology: Items 16, 21, and 22 of this survey were used as a measure of fundamental knowledge (DF educational outcome 1). Item 36 was used as a measure of the ability to frame and resolve ill-defined problems (DF educational outcome 2). Item 39 was used as a measure of intellectual curiosity (DF educational outcome 6). This information was gathered for all core courses at USAFA. These questions were compared between the divisions and with CWI. These questions were also compared across the 100, 200, and 300/400 course level.

Assessment of the Social Sciences Division: Overall the Social Sciences Division is at or near the top of most of these ratings. However, because of the tight grouping of these values, except for low ratings of the Engineering Division, not much comparative information is provided.

DF Educational Outcome 1: For Question 16 "intellectual challenge," all divisions are high (4.5-5.0/6.0) except CWI, which is much lower, and the Basic Sciences Division, which is slightly higher in 3/400 level courses. For Question 21 "relevance of course content," the Social Sciences Division is slightly higher than the other divisions except during the 3/400 level courses, when the Engineering Division is extremely low. For Question 22 "amount learned," the Social Sciences Division is highest for 100 level courses, about equal to the other divisions for 200 level courses and in a cluster at the top for 3/400 level courses. The Engineering Division is again low for 3/400 level courses. Taken together these three questions (16, 21, 22) were the best measure of DF Educational Outcome 1 "fundamental knowledge" from the Dean's End of Semester Critique. The Social Sciences Division is at or near the top in all comparisons for this educational outcome.

DF Educational Outcome 2: For Question 36 "problems without approved solutions," all divisions were again high (all above 4.5/6.0 and most above 5.0/6.0). The Social Sciences Division was the highest across 100, 200, and 3/400 level courses. The largest difference was compared with the moderately lower Engineering Division.

DF Educational Outcome 3: For Question 39 "read more," there was considerably more variance with a range between approximately 4.4 to 5.5. The Social Sciences Division was at or near the top across the 100, 200, and 3/400 level courses. The largest difference was the moderately lower Engineering Division in the 3/400 level courses. Also noteworthy was CWI's very high score on the 3/400 courses.

Conclusion: Taken together a clear pattern emerges. The Social Sciences Division is at or near the top of all these questions. It appears that from the students' perspectives, as reported on the Dean's end of semester critique, all divisions, especially the Social Sciences Division, are accomplishing DF Educational Outcomes 1, 2, and 6.

DF Educational Outcomes Coverage Survey

Course directors and instructors rate the level of coverage that their course is designed to provide for each of the seven educational outcomes. The range of the rating scale is from 0 (no coverage) to 3 (heavy coverage).

Methodology: The individual ratings were grouped by course, then by department, then by division. Averages for each educational outcome were determined for each division. These averages were then compared with the faculty-wide averages. Particular attention was paid to three educational outcomes: Fundamental Knowledge, Ill-defined Problems, and Intellectual Curiosity.

Assessment of the Social Sciences Division:

The averages for DF and the Social Sciences Division are summarized in Table 1. This provides a cursory description of how those teaching the Social Sciences Division core courses view their contribution to the seven DF educational outcomes. More information was generated by looking at the Social Sciences Division with respect to each of the other divisions. A plot of each division's average for each of the outcomes is shown in Figure 1. The Social Sciences Division rates higher than the DF average in each of three outcomes of interest. Of particular note, on outcome 2 (framing and resolving ill-defined problems), the

Social Sciences Division rates the highest. For this outcome, the data reinforces the expectation that the "hard" divisions tend to have more well-defined problems, while the "soft" divisions tend to focus on ill-defined problems. (NOTE: the *statement* of this outcome may influence how each division assigns their ratings. In basic sciences, though a problem may be well-defined, the path towards deriving the solution may, in fact, provide great latitude for creative thinking. Perhaps a more appropriate description of the outcome would be "solving problems with ill-defined solutions.") A similar trend exists for intellectual curiosity (DF Educational Outcome 6). The Social Sciences Division, along with the other "soft" division, rate significantly higher than the two "hard" divisions. In summary, the instructors and course directors for the core courses in the Social Sciences Division tend to emphasize fundamental knowledge, framing and resolving ill-defined problems, and intellectual curiosity more than the typical course at the Academy.

	1*	2*	3	4	5	6*	7
DF	2.53	1.96	2.02	1.86	2.16	1.93	1.93
SS Div	2.60	2.53	2.35	1.90	2.02	2.26	2.23

Table 1: Coverage Rating Averages
(* denotes DF Educational Outcome of Interest)

Course directors assess the grading policies of their course. Ratings include the percentage of the total course points given to: final exam, graded reviews, multiple choice, essays, computational work, performance, out-of-class work, collaborative work, assignments for which re-accomplishment is
allowed, and student choice in
graded assignments.

Methodology: From the course director input, DF averages for each category were calculated. Course director input was consolidated by division and division averages for each category were calculated.

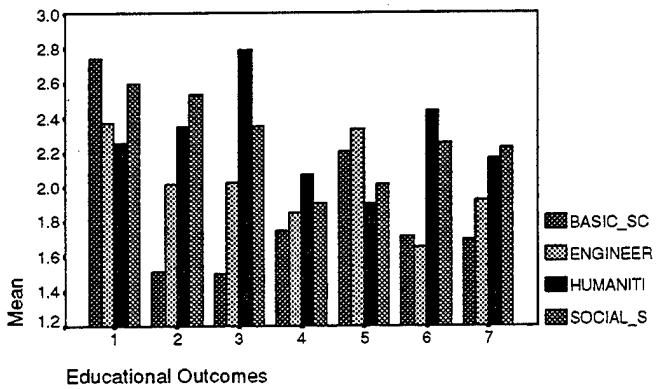


Figure 1: Plot of Average Ratings by Division for Each Outcome

Assessment of the Social Sciences Division: Combined with the coverage ratings from the previous instrument, these data indicate whether the course directors assign points accordingly. The percentages of final grades from multiple choice, essays, performance, and computation (which sum to 100%) for the Social Sciences Division are as follows: 45.4%, 43.6%, 10.6%, and .33%, respectively. Though the high percentage of multiple choice seems counter to the high coverage ratings for "Ill-defined Problems," it is important to recognize that this division also rates "Fundamental Knowledge" highly. One can speculate that the Basic Sciences and Engineering Divisions grade "Fundamental Knowledge" through a combination of multiple choice and computational problems, while the Social Sciences Division may seek to accomplish this through multiple choice problems and essays.

The third outcome, "Intellectual Curiosity," is difficult to measure from the categories defined in this instrument. Giving students some degree of assignment choice is a method of stimulating their interest and may align with outcome 2. The Social Sciences Division ranks second among all divisions with an average of 7.4% of the course grade subject to student choice of assignments. However, this figure is below the DF average of 12.1%. Though the DF average is skewed by the extremely high average of the Humanities Division (40.75%), looking back upon the *coverage ratings* (see previous instrument) for outcome 6, we would expect similarities between the Social Sciences and Humanities. In fact, we see a dissimilar allocation of points for "student choice." *If the student choice percentage is a good indicator of a course's attempt to develop intellectual curiosity*, then there is an inconsistency between the Social Sciences Division's *desire* to emphasize intellectual curiosity (as measured by the coverage ratings) and their *actual* emphasis of intellectual curiosity (as measured by grade percentage).

COMPOSITE RATINGS FOR SOCIAL SCIENCES

DIVISION CORE COURSES

The Process

For each of the three outcomes, courses were rated within the Social Sciences Division recognizing that the division scores were higher in these categories than the other divisions in DF. The group gave a 5 rating knowing that a particular course with this rating may rate higher when compared to DF. We used a relative scale of 1 to 7. We did not apply the descriptions suggested in the Phase II Charter (i.e., a 5 did not mean that certain policies or practices may be particularly noteworthy but consistency across sections or course segments could improve). A 5 meant an above average rating but not as good as a 6 or 7.

The group debated the role of 200 level courses. If the course laid the groundwork for high attainment of educational outcomes in follow-on courses (particularly the integrating fundamental knowledge outcome), should it be rated high even if it didn't seem to contribute to ill-defined problems? We chose to rate it low even though we understood that for most outcomes, 200 level courses should rate lower than 300 level courses which, ideally, rate lower than 400 level courses. For integrating fundamental knowledge, cadets gave high ratings to the 200 level courses that built on high school knowledge. For example, students

probably rated PS 211 high in integrating fundamental knowledge because of the exposure they had in high school to courses like Civics and Social Studies.

Courses were rated within the division based on cadet responses on the Course Assessment Survey. Question 12 was examined for relevancy to the intellectual curiosity outcome; questions 18-21 for integrated knowledge; and question 17 for framing and resolving ill-defined problems. Major breakpoints in the rating correspond to different final ratings. For a sanity check, these results were checked against the instructor ratings of courses for the three outcomes. Although looked at initially, the percentage of final grade metrics was not a significant part of the final outcome. When the framing and resolving rating was assigned, the group reviewed the percentage of multiple choice versus essay, and the percentage of writing required in the course. After rating the courses we looked for gaps to determine how we rated them based on the 7 point scale. This process was done without regard to how a prerequisite course may have affected the following course.

The Ratings

The ratings were also accomplished without considering the DF ratings (though these were reviewed after the ratings were assigned). The group focused primarily on the relative standings within the division. The course instructor assessments for each outcome were compared to the average rating by division, and to the DF-wide overall averages.

Course	Integrated Knowledge	Intellectual Curiosity	Framing and Resolving
BS 110	5	6	6.5
BS 310	6	6	7
Econ 221	5	5	6
Econ 310	7	6	6
Law 320	6	6	6.5
Law 420	7	7	7
Mgt 210	6	5	6
Pol Sci 211	7	7	6
Pol Sci 212	6	6	6

CONCLUSIONS

Outcome 1: *Fundamental Knowledge*

The primary conclusion of the committee was that each core course is contributing to the overall integration of knowledge within the division. As noted, the committee reviewed the data collected from the Course Assessment Survey and noted that the students were able to discern a higher level of integration in the upper level courses than in the lower level courses. In all but one case, cadets rated the "building block" course (BS 110, Econ 221, Law 320) lower on integration than its corresponding upper level course (BS 310, Econ 310, Law 420). The instructor assessments indicate that the emphasis on "fundamental knowledge" is higher in the "building block courses." That is consistent with the conclusion that instructors in

those upper level courses assume a certain level of fundamental knowledge about the subject matter, and use that knowledge for a more sophisticated presentation and analysis. The committee concluded from the data that the "building block" courses within the division are performing their designed function of providing the fundamental knowledge necessary for integration in later courses.

The anomaly to this overall design was PS 211 and PS 212. It was the committee's conclusion that PS 211 benefited from high school courses which, in effect, provided the basic "building blocks" in Civics and American government which are absent in PS 212. Nevertheless, PS 212 rated relatively high within the division on this outcome.

Outcome 2: ***Framing and Resolving Ill-defined Problems***

The committee examined Question 17 from the Course Assessment Survey and also examined the data regarding the type of testing involved in each core course. The committee looked for consistency between data rating a course high and the testing techniques utilized (that is, a higher level of emphasis on "non-objective" kinds of examination tools). With the notable exception of BS 310, which did rely on a high level of objective questions in its examinations, there was some correlation between the ratings and the percentage of an examination dedicated to essay or "non-objective" testing. The committee concluded that as a division, all core courses provide a high level of framing and resolving ill-defined problems and that the cadets recognize this divisional emphasis.

Outcome 6: ***Intellectual Curiosity***

The committee decided to reject questions from the survey dealing with outside reading, but looked at the overall statistical ratings, and in particular, the results of Question 39 of the Dean's End of Semester Critique. The committee concluded that the greater the perceived relevancy of a course to a student's career and future life experience, the higher the rating. Accordingly, the social science division rated very high. The "High" rating by courses in law and politics (Law 420, PS 211, Law 320), in particular, were viewed by the committee as supporting this conclusion. Likewise, the data indicated a higher interest in the field of economics by the more senior students, again supporting the conclusion that as a subject matter's relevancy became more obvious, its rating increased (Econ 310 rated 3/9 overall in this category, while Econ 221 was 8/9). In summary, the committee concluded that each core course fostered a desire within the student to learn more about its subject matter and that students clearly appreciated the relevance of the courses to their intellectual development as well as their careers.

Social Sciences Division Commentary & Critique

1) ***What has this report revealed about the three educational outcomes from the perspective of the Social Sciences Division?***

The Social Sciences Divisional Team considered all three educational outcomes to be important but claimed the ability to frame and resolve ill-defined problems as a particular divisional specialty. Society and its institutions are inherently fuzzy concepts and the application

of scientific protocols is seldom straight forward. "Plug and chug" approaches to problem solving which may work well with more concrete problems are seldom appropriate or effective when applied to social sciences. The team also suggested that the social sciences had a relative advantage concerning students' intellectual curiosity. This may reflect a collective personal bias on the part of committee members, but all agreed that social science topics were more interesting than those in other academic divisions. Although developing students' integrated fundamental knowledge was not seen to be unimportant, the team felt that from a social sciences perspective, this outcome should not be elevated above either of the other two outcomes. The team also asserted that the higher the level of the course, the higher the score for integrated knowledge should be; their report contains an argument for why this should be so but no independent evidence other than the fact that they rated upper division courses more positively.

All courses within the division were rated between 5 and 7 for their contributions to integrated knowledge; between 6 and 7 for their contributions to cadets' ability to frame and resolve ill-defined problems; and again between 5 and 7 for the courses' contributions to intellectual curiosity. The team seemed to place a fair amount of reliance on the fact that most of the instruments indicated that social sciences courses and faculty scored slightly higher than the faculty averages to support the conclusion that there were few if any problems. A minor concern regarding the expressed commitment to increasing students' ability to frame and resolve ill-defined problems and the relatively high use of multiple choice tests as an evaluative tool. However, this concern did not prevent the team from rating all courses as either 6s or 7s for their contributions to this outcome. Similarly, there was a minor concern voiced about the apparent inconsistency between a commitment to increasing intellectual curiosity and the relatively few cadet choices allowed in assignments. Again, this concern did not appear to be translated to lower assessed contribution ratings for any particular course.

2) ***What has the report told us about the contributions of specific core courses in the Social Sciences Division to the educational outcomes?***

Due to process adjustments noted below, all social sciences core courses were rated relatively highly (at least "above average") on all three outcomes. The top overall contributor was seen to be Law 420 which was awarded perfect 7s on all 3 educational outcomes followed closely by Political Science 211 which received 7s on knowledge and intellectual curiosity but only a 6 on framing and resolving ill-defined problems. Several courses received relatively low scores for their contributions: Economics 221 (5,5 & 6) and Management 210 (6, 5, & 6) were the two lowest followed closely by Behavioral Sciences 110 and Political Sciences 212. However, the abandonment of the common rating scale and the apparent propensity to make assumptions then assign ratings to match, render these course contribution scores as suspect at best.

3) ***What can be learned from the Social Sciences Division Assessment Report concerning the assessment process and practice in general?***

This divisional assessment team worked together to achieve consensus and documented their process in sufficient detail to allow post hoc analysis. These are both strengths. However, it must be noted that the team also made several assessment decisions which cast doubt on the

comparability and even validity of the ratings assigned. The team arbitrarily abandoned the agreed upon assessment rubric agreed to by the other teams (the 1 to 7 scale previously presented) and adopted a truncated ordinal scale. This choice seemed to have been based on the somewhat dubious assumption that since the social science divisional average contributions were higher than the faculty averages on most measures, all social sciences core courses should receive "above average" ratings (of 5 or higher). Ratings of 6 or 7 were used to connote even better contributions. Although this approach greatly attenuates the comparability with scores provided by other teams, it would have at least retained the ordinal validity of scores. The team's further assumption that higher level courses "should" receive higher ratings for integrated knowledge introduced a systematic bias in favor of upper division courses and against 100 and 200 level courses. One symptom of these anomalies is the fact that two of the assessments differed from those provided by the horizontal assessment team by 2.0 (Econ 310's contribution to knowledge and BehSci 310's contribution to ill-defined problem resolution ability). Also the overall average rating disparity between the Social Science Division team's ratings of their courses and those provided by the horizontal teams was .8 (compared to .1 for Basic Sciences and .3 for Humanities). All in all, this assessment report's primary contribution is to show the consequences of deviating from standard protocols and allowing rationale to replace evidence.

Chapter 6

Engineering Division Report

CHAPTER SIX - ENGINEERING DIVISION REPORT

Dr. Daniel J. Pack--Department of Electrical Engineering
Maj Michael K. Fabian--Department of Aeronautics
Capt Nick Seward--Department of Astronautics
Maj Blaise A. Horban--Department of Engineering Mechanics
Major William A. Kitch--Department of Civil Engineering

The principle objective of the working group for Spring '96 was to assess the following three educational outcomes which deal with knowledge, skills, and attitudes of cadets upon completing each of engineering division core course: *Integrated Knowledge, Framing and Resolving Ill-Defined Problems, and Intellectual Curiosity.* The findings in this report are based on results from end of semester course critiques (Fall\95 and Spring\96), core course assessment questionnaire for cadets (Spring\96), curriculum review questions for instructors (Spring\96), and curriculum review questions for course directors (Spring\96). The numerical values associated with three educational outcomes in this report should **not** be considered as a sole source to evaluate the outcomes. Furthermore, validity of some of the data is questionable due to some answers being formulated without clear understanding of terminology. Throughout this report each educational outcome is rated for each course with a numerical number ranging 1 through 7, where 1 denotes unsatisfactory level of contribution, 7 is for best possible contribution.

INTEGRATED KNOWLEDGE

Most of engineering division core courses address this particular educational outcome effectively as shown in this report. In particular, freshman and sophomore level courses emphasize this outcome more than two other outcomes we are considering. According to instructors for those courses, this emphasis is natural since those classes are used to lay ground work for cadets to move on to greater academic challenges in their junior and senior years at the Academy. In this section we illustrate how each of the engineering division core courses promotes this educational outcome.

Engr 110 Introduction to Engineering

In terms of integrating fundamental knowledge, this course rates a 6 for this educational outcome. Given the broad nature of the design problems presented and the need to assimilate various aspects of the design process, the students must integrate aspects of basic physics and engineering to achieve their design goals. This is one of the most important aspects of the course according to the course director survey.

Engr Mech 120 Fundamentals of Mechanics

Engineering Mechanics 120 (EM120) is taught to first semester sophomores or second semester freshman during the Fall or Spring semester, respectively. The assessment tool was administered to second semester freshmen. The objectives of EM120 are to learn fundamental knowledge

about static equilibrium of structures and strength of materials. This knowledge will be helpful for future Air Force officers in understanding the relationship between forces acting on aerospace structures and the material response to those forces. This knowledge will be helpful in understanding why structures fail and to a lesser degree how to prevent those failures. The course draws primarily upon fundamental math and physics knowledge. The purpose of EM120 is to expand the breadth of fundamental knowledge cadets possess. Predictably the cadet responses to questions 18-21 which are primarily intended to address the first educational outcome are above the averages for overall core course means. The largest delta occurs with "uses knowledge I have learned in another course." There appears to be a positive 0.50 difference for this item in the overall core course average versus EM120. This is primarily due to the integration of math and physics principles in solving EM120 problems and in understanding the derivation of many of the guiding principles in EM120. The only question where there is not a significant difference is with the responses to question 21. The 3.55 scored here still indicates less than agreement (4) with the question but better than neutral (3). All responses to this series of questions had a large standard deviation given the limited number of responses. This large standard deviation indicates a fairly strong disagreement as to how much is both learned and integrated with other courses. Overall rating: 6.

Aero 215 Fundamentals of Aeronautics

Aero 215 rates a 6 on the scale for this issue. The cadets learn the terminology and basic equations of aeronautics and apply these concepts to wings and aircraft to determine performance parameters relevant to aeronautics in the later portion of the course. The instructors and course director placed the highest emphasis in course content on this issue. The cadets felt this course had more relevant content than any engineering core course. Similar cadet feedback occurred for "amount learned" in the course.

El Engr 215 Electrical Signals and Systems

EE215 is a sophomore level core course taken by all cadets except those who major in Physics, Computer Science, Electrical Engineering, Astronautical Engineering, Civil Engineering, Engineering Mechanics, and Mechanical Engineering. The course is designed to teach cadets fundamental knowledge on signal analysis and electronic system design. Main emphasis of this course coincides with this educational outcome. Throughout the course, cadets will be exposed to fundamental concepts in electrical engineering. Cadets learn how signals are represented mathematically, in the time domain, as well as in the frequency domain, and how those signals are used to communicate between two physically different stations. Cadets also learn about data representation and manipulation in digital systems, and how to analyze circuits with passive and active electronic devices. The course includes four separate labs to help cadets "try out" what they learned in the classroom. The instructor survey illustrates that this educational outcome is a main focus of this course. The average numerical value from instructors was 2.6/3.0. Cadets seem to agree with instructors, according to both the end-of-semester critique and the course assessment questionnaire. The end of course critique questions related to the outcome received an average of approximately 4.2/6.0, whereas the course assessment questionnaire resulted in average value ranging 2.87 to 3.33 on a 5.0 scale. For the course assessment questionnaire, question numbers 18 through 21 are used. The number of graded exams, quizzes, labs, and

worksheets demonstrate the emphasis of this course on the fundamental knowledge for this particular outcome. The overall rating of the outcome contributed by EE215 is 6.0/7.0.

El Engr 231 Electrical Circuits and Systems I

EE 231 is a sophomore level core substitute course taken by cadets majoring Physics, Computer Science, and all engineering disciplines except Aeronautics. The main emphasis for this course is to introduce tools for circuit analysis and to teach concepts for electronic system design. This course plays an important part of this particular educational outcome for cadets who declared the aforementioned specialties. Obtaining knowledge of electronic fundamentals and circuit analysis techniques by the majors is required for cadets to be successful in their fields. Physics majors often analyze problems in frequency domain, use electronic filters, and deal with basic electric variables of energy and power. Computer Science majors need to understand electronics for applications in hardware. All the above majors learn fundamentals on communication systems. According to the results of curriculum review questions for instructors, instructors agree this category of the outcome plays an important role for this course (3.0/4.0 where 4.0 represents heavy emphasis). Cadet survey agree with instructors on this educational outcome as the results of the related questions show average values ranging from 3.51 to 3.74 out of 5.0 where numerical value 5.0 stands for "strongly agree." The average value for the end of course critique related questions resulted approximately 4.3 out of 6.0. Also, the course director's assignment of graded materials on this educational outcome concurs with the findings from the instructor and cadet surveys. An overall rating for this category contributed by EE231 is 6.0/7.0.

Civ Engr 310 Air Base Design and Performance

Civ Engr 310 is unusual among Engineering Division Core courses in that it integrates knowledge from courses outside the Engineering and Basic Sciences Divisions. In particular, Civ Engr 310 draws from history, political science, and military arts and science courses. Cadets are required to integrate knowledge across many divergent disciplines to solve the given problems. Overall rating: 6.0

Engr 310 Energy Systems

Engr 310 rates a 5 on integrating knowledge. The first half of the course involve the fundamental laws of thermodynamics, substance properties and appropriate equations for substances. The second half of the course involves the use of these relationships for various energy systems for power generation, environmental heating and cooling, and jet propulsion. This was the highest rated area for the course content from the instructor and course director survey.

Engr 311 Electrical Power System

Engr 311 is a junior level core substitute course for Eng 310 for cadets majoring in Civil Engineering, Computer Science, Physics, and Electrical Engineering. The main emphasis of the course is to teach cadets how mechanical energy is converted to electrical energy and how the energy is distributed among users. The course is designed to teach cadets the principles of energy conversion to electrical power systems. Cadets learn basic knowledge on motors, rotors, DC motors, AC motors, and power supplies. Also, the conversion between AC to DC as well as 3-phase circuits are studied in the course. The underlying theme for all these subjects is the basic thermodynamics principle which says the level of energy stays the same, only its form changes.

Cadets see this course as one whose emphasis is on this particular educational outcome. The end-of course critique shows average numerical values of 4.3 and 4.4 out of 6.0 for questions 21 and 22, respectively. The course assessment questionnaire results show values ranging from 2.76 to 3.32 out of 5.0 with standard deviation over 1.0 on average. Questions used for the above data are numbers 21 and 22 from the end of course critique, and numbers 18 through 21 on the course assessment questionnaire. The overall rating of this course on integrated knowledge is 6.0.

Astro 320 Introduction to Astronautics for the Engineer and Scientist

Astro 320 is the core Astronautics course for technical majors. Astro 320 is very strong in providing fundamental knowledge to cadets about orbital motion and space operations, but it also provides a solid foundational knowledge in FORTRAN computer programming. Astro 320 covers all of the same topics as Astro 410 (orbital motion, orbital maneuvering, orbit prediction, reentry, and space mission design), but cadets are also taught how to use structured programming techniques to solve complex orbit prediction problems. Overall rating: 6.0.

Astro 410 Introduction to Astronautics

Astro 410 is the core Astrodynamics course for non-technical majors. Astro 410 provides a solid fundamental knowledge base concerning orbital motion, orbital maneuvers, orbit prediction, reentry, and space mission design. Cadets are given a basic knowledge of the advantages and limitations of conducting space operations. The course focuses more on the fundamental concepts of astronautics, complemented by a review of current technologies and operational systems. Overall rating: 6.0.

Engr 410 Engineering Systems Design

Engr 410 is the core capstone engineering design course. Engr 410 provides minimal fundamental knowledge. During the first seven lessons, the Department of Defense Systems Acquisition Process is presented and discussed, including examples of successes and failures of implementing this process by the different armed services in the past. Beyond this material about the Systems Acquisition Process, cadets are expected to enter the course with the fundamental knowledge required to successfully accomplish their class project. Engr 410 gives cadets the opportunity to apply their solid fundamental knowledge base that they have acquired during their first 3 years at the Academy and work as a team to accomplish their project. Overall rating: 4.0.

FRAMING AND RESOLVING ILL-DEFINED PROBLEMS

Junior and senior level courses in the engineering division promote this educational outcome effectively. Freshman and sophomore courses are working on this educational outcome with some success. In this section, each core engineering course is addressed with respect to this outcome.

Engr 110 Introduction to Engineering

This is a new course and it rated a 7 for this educational outcome. This entry level course is designed to expose cadets to the broad spectrum of issues in design projects. Part of last semester's course was planning a mission to Mars. On the cadet course critiques, this course was the highest rated course in the engineering division for having problems without approved solutions. On the EOAWG student survey, the two questions with the highest cadets responses

were "This course lends itself to ill-defined problems" and "I frequently found myself actively engaged in thinking about difficult questions for which we still need to find answers." In this course, material is presented by instructors on an "as needed" basis as the flow of the design uncovers new engineering issues.

Engr Mech 120 Fundamentals of Mechanics

The questions attempting to address the second educational outcome, resolving ill-defined problems, present a more mixed basis for discussion. Generally, the responses reflects marginal agreement that this outcome is being addressed. Critique questions 7, 9 and 16 come closest to addressing this outcome. In all three cases, EM120 students are near agreement (response of 4) that the course is helping them to resolve ill-defined problems. The remainder of the questions dealing with the second outcome are more ambiguous. Although, EM120 remains near the average for all core courses. The cadet response average indicates an opinion that is between neutral and agreement and with a large standard deviation. Overall rating: 5.

Aero 215 Fundamentals of Aeronautics

This course would rate a 4 on this educational outcome. The emphasis in Aero 215 is on fundamental knowledge of basic fluid mechanics and aircraft performance. It rated as being middle of the pack on the cadet surveys for "problems without approved solutions." It was rated low by the instructors and the course director due to the lack of a design project which is featured in most other aeronautics courses. Spreadsheet exercises with associated questioned represented the largest portion of the ill-defined problem aspect of the course. An optional balsa aircraft project provides much of the ill-defined aspect of the course, where the cadets must build, analyze and fly their aircraft.

El Engr 215 Electrical Signals and Systems

The nature of the course dictates the limited emphasis on framing and resolving ill-defined problems. Cadets design transducer interface where cadets need to devise a numerical gain and a bias value to match the input to an interface with the output in order to make the interface function. Cadets also need to design and determine parameters involved in analog-to-digital converter. Instructor survey results show this outcome does not go hand-in-hand with EE215 because of the course nature--designed to expand the scope of cadets' fundamental knowledge on the Electrical Engineering subjects. The average numerical number given by the instructors is 1.0/4.0. Cadet surveys resulted in a wide range of numerical values. For the course assessment questionnaire, the average values range from 2.46 to 3.56 out of 5.0 (questions 7,9,10, 13, 16, and 17), and the standard deviation was on average more than 1.0. Answers to the end of course critique question(question number 36) resulted in average numerical value 4.7/6.0. The results show that the definition of ill-defined problems was not clear to cadets. The overall rating of EE215 on this educational outcome is 5.0/7.0.

El Engr 231 Electrical Circuits and Systems I

Several design activities conducted in this course fall into this educational outcome category: transducer interface, op amp, filter, and communication systems. For these problems there is no single correct answer, and cadets are required to make certain assumptions to arrive at their answers. During the past year, more emphasis was given on this category of the outcome.

Instructor survey gives numerical value 2.0/4.0 for this category, where again 4.0 represents heavy emphasis. Only one of the end of course critique questions address the particular educational outcome and the average value for the result was 4.8 out of 6.0. Cadet surveys show fairly high average value to support that there are ample amount of materials that are associated with this educational outcome. The average values range from 3.05 to 4.01/5.0. An overall rating for this category for EE231 is 6.0/7.0.

Civ Engr 310 Air Base Design and Performance

This course is targeted directly at this particular educational outcome. The course is structured around two major projects which require students directly attack large ill-defined problems. The cadets recognize this. The mean score on the EWOG course assessment question number 17, "this course lends itself to ill-defined problems," Civ Engr 310 earned the second highest score of any Engineering Division courses (3.62). Overall rating: 6.0.

Engr 310 Energy Systems

In this area, Engr 310 rates a 5 for this educational outcome. On the recent EOAWG survey the cadets rated this area the lowest. Unfortunately, the survey was conducted prior to the design project which occurs near the end of the course when the cadets finally have enough knowledge to tackle a project. In recent years this aspect of the course has been reduced due to cadet time issues and new computer software experiments. Most of the course time is devoted to teaching the cadets the tools they will need to analyze thermodynamic systems, and their knowledge level limits the difficulty of the design problems.

Engr 311 Electrical Power System

A special activity each cadet participates in is the final project which is to design a power plant which generates enough power and distributes it to an entire Air Force base. Cadets need to make decision changing assumptions on the use of DC motors, DC generators, and AC motors. They need to consider cost versus efficiency on their designs throughout the project. Initial response of this project from cadets is that the project gave them a chance to use what they learned in class and the project was challenging and worthwhile. The overall rating of the course for this outcome is 6.0/7.0.

Astro 320 Introduction to Astronautics for the Engineer and Scientist

Astro 320 is not strong in presenting ill-defined problems. The course readings, assigned homework problems and computer projects are very straight forward. The general algorithms for the computer programs are discussed and presented to the cadets from a big picture perspective. The main challenge for the cadets is to implement the algorithms successfully using their understanding of correct computer syntax. Overall rating: 4.0.

Astro 410 Introduction to Astronautics

This course does not emphasize ill-defined problems. The course material, assigned homework problems, and graded homework problems are very well quantified and bounded. The only exception to this would be the end of course design project, which is worth 10% of the overall course grade. This project requires cadets to make decisions about which spacecraft payloads and subsystem components should be included on their spacecraft and what orbit it should be placed

in so that it can accomplish a given mission. The cadets are given numerous missions to chose from. An interesting aspect of this project is that cadets often do significant mission analysis of each of the possible missions just to determine which mission would be the easiest to design! Overall rating: 4.0.

Engr 410 Engineering Systems Design

Engr 410 presents a great opportunity for cadets to solve an ill-defined problem. Projects are selected by course instructors before the beginning of the semester. Instructors are directed by the course director to define the problem specifically enough so that the end product meets the actual need, but often requirements are written that cadets soon discover must be clarified or even challenged altogether for their validity. It is interesting to note that on the EOAWG course assessment, Engr 410 students agreed more strongly (mean 4.03) than any other statement on the assessment, that Engr 410 "lends itself to ill-defined problems." The rating for Engr 410 for this educational outcome is excellent, 7.0.

INTELLECTUAL CURIOSITY

The task of coming up with tools to measure this particular educational outcome was most challenging. The engineering EOAWG group believes that more study should be done to come up with reliable means to evaluate this particular educational outcome. In this section, we show how each engineering core course is addressing this educational outcome.

Engr 110 Introduction to Engineering

Although "Intellectual Curiosity" is difficult to define, this course would probably rate a 7 for this educational outcome. In terms of intellectual challenge, the students rated this course the highest of the Aeronautics Department core courses and one of the two highest in the engineering core. It was second behind Astro 320 (a course for engineers only) in terms of motivating the students to read more in the subject area. Surprisingly, the course director rated intellectual curiosity as receiving relatively little emphasis. From the course director's viewpoint, it may be an issue of the meaning of "intellectual curiosity."

Engr Mech 120 Fundamentals of Mechanics

The final outcome, intellectually curiosity, is more difficult to analyze due to limited data. Questions 8 and 12 provide the best basis for analysis. Responses to both questions showed that EM120 is in the middle regarding addressing this outcome. Both questions indicate better than neutral responses with a large standard deviation. Overall rating: 5.

Attempting to compare this cadet assessment data to the end-of-course critique data raises some questions. First, based on the data dealing with knowledge gained or used in the cadet assessment tool, cadets learn more than they reflect on the student critiques. In fact, assessment tool questions 18-21 dealing with fundamental knowledge, EM120 does much better in this area than the critique data indicate. This statement is based on the responses to question 22 of the end-of-course critique. Additionally, based on the response to the end-of-course critique question 36, 'this course improved my ability to deal with problems that don't have an approved solution'. EM120 received a mark of 4.7 (between good (4) and very good (5)). However, on the educational outcomes assessment tool (questions 7 & 9) EM120 did not even receive an agree

response (4) but had an average for both questions around 3.75. Additionally, comparing the data for EM120 on the end-of-course critiques and the assessment tool with the other division's data shows a discontinuity. The assessment tool data indicate EM120 is better than average in meeting two educational outcomes tested directly. However, the end-of course data indicate that EM120 is less well regarded than courses in other divisions. This is particularly true for intellectual curiosity. The discrepancy is greatest between question 39 on the end-of-course critique and the assessment tool. These questions may be affected by when the assessment tool was administered during the semester.

The critique was given to all freshmen in EM120 just prior to recognition training. Given that their primary concern for those few remaining days of their fourth year was anything but academics, the reliability of all data provided by the freshmen has to be questioned. The disparity between the data on the assessment tool and the end-of-course critique can be explained in part by recognition training.

Aero 215 Fundamentals of Aeronautics

In terms of "Intellectual Curiosity," this course rates a 5. Based on the entering freshman desires for academic majors, a large majority of cadets are interested in aeronautics and airplanes. The cadets seem to enjoy the Aero 215 course and subject material. The course rated third in the engineering division for the cadets wanting to read more.

El Engr 215 Electrical Signals and Systems

This is one of the outcomes that is very difficult to measure. At this point, the only data that we can base our opinion on are the surveys mentioned in the beginning of this report. Instructors gave numerical value 1.0/3.0 to this outcome. Answers to the end of course critique questions 39 and 16 were 4.5 and 4.3 out of 6.0, respectively. Cadet course assessment questionnaire results showed numerical values 2.75 and 3.15 out of 5.0 (question number 8 and 12). The standard deviation for the course assessment questionnaire result was over 1.0. The overall rating for EE215 on this educational outcome is 5.0/7.0.

El Engr 231 Electrical Circuits and Systems I

This is one of the most difficult educational outcomes to assess. One indication of cadet curiosity was noted by the course director after labs have finished. While non-curious cadets put everything away and wait to leave, curious cadets "play" with the equipment to observe other non-required experimental results. Instructors view this course as having a small emphasis on the educational outcome and gave numerical value of 1.0/3.0. The end of course critique contains two questions related to this educational outcome. The results shows numerical average values 4.6 and 4.8 out of 6.0. Cadet surveys give 3.24 and 3.18 on related questions (specific question numbers: 8 and 12). Although the cadet curriculum survey and the end of course critique agrees on this category, there needs to be more sound means to measure this outcome, rather than relying on solely on cadet responses. An overall rating for this category of the educational outcome for EE231 is 5.0/7.0.

Civ Engr 310 Air Base Design and Performance

Without a satisfactory assessment tool it is difficult to evaluate this outcome. Anecdotal evidence from student reaction papers, and written comments on end-of-course critiques, indicate that

many students do find that the course has stimulated their curiosity. However, on the few assessment tools available, such as Question 16 from the end-of-course critique dealing with the "intellectual challenge," Civ Engr 310 rates as one of the two lowest courses in the Engineering Division. Overall rating: 5.0.

Engr 310 Energy Systems

In this area, the course rates a 4. Some cadets get very excited about the course material on automobile engines and turbo/super-charging, while other cadets are motivated by the portion on jet engines. Unfortunately, the formal and math-intensive nature of thermodynamics limits the excitement for many cadets.

Engr 311 Electrical Power System

To further initiate cadet curiosity, cadets are given articles from professional journals to read. The article is not discussed in class but is included in the test materials. Some cadets have come to instructors with questions which indicate cadets are curious about the subject matter. A reliable assessment tool is needed. The end of semester critique questions resulted in average numerical numbers 4.7 and 4.8 out of 6.0 for questions 16 and 39. The course assessment questionnaire results show numerical values 3.37 and 2.91 out of 5.0. The overall rating of this course on the educational outcome is 5.0/7.0.

Astro 320 Introduction to Astronautics for the Engineer and Scientist

This course stimulates intellectual curiosity in cadets because many cadets find space operations and space itself to be fascinating topics. Cadets often bring in articles from outside sources that relate to the course material, or ask questions about current space missions that relate directly to the course material. Overall rating: 6.0.

Astro 410 Introduction to Astronautics

The course covers topics that many cadets find interesting and often results in cadets bringing in outside sources about the subject matter, or asking questions about current space systems and missions that are closely related to the course material. Overall rating: 6.0.

Engr 410 Engineering Systems Design

This course challenges the imagination of the cadets to think of their own solutions to solve their given problem. Early in the course, during the Ideation Phase, cadets are encouraged to consider any and every possible solution to their projects. Overall rating: 6.0.

Engineering Division Commentary & Critique

1) ***What has this report revealed about the three educational outcomes from the perspective of the Engineering Division?***

The Engineering Division Assessment Team showed how each course contributed to each of the three educational outcomes: integrated knowledge, ill-defined problem solving and intellectual curiosity. It appears as though courses were assessed independently (i.e., by each respective departmental representative) because the quality and quantity of support for the

assessment ratings for each outcome varied considerably from course to course. There seemed to be a common tendency to consider the neutral point on each instrument as indicative of "average"; since few scores yielded below neutral ratings, no contributions were rated as being deficient and very few as only average. There was also a strong inclination in some ratings to rely on what was intended or what was taught rather than what was learned or experienced by students. Several of the assessments contain very thorough support in terms of curricular content.

Although integrated knowledge seemed to be the best understood and most strongly endorsed outcome, the team appeared to have some difficulty in distinguishing between various levels of contributions by the ten engineering courses considered (eight courses received a rating of "6"; one received a "5" and another received a "4"). In contrast to the Social Sciences Division Team, however, the Engineering Division Team assumed that courses earlier in the curriculum were more oriented toward the development of integrated knowledge; they suggested that junior and senior level courses should stress application and problem solving rather than knowledge acquisition.

Upper level (3-400) core engineering courses were seen as being the strongest contributors to cadets' ability to frame and resolve ill-defined problems. The one very notable exception to this is the new Engineering 110 course which uses a problem solving orientation to introduce cadets to general engineering principles as well as an overview of the various types of engineering and alternative approaches to solving real problems. The 7 rating this course received suggests that, at least for the Engineering Division, this challenging interdisciplinary course represents great potential for future development.

The team seemed to encounter the greatest difficulty in assessing intellectual curiosity; there appeared to be a great deal of diversity of opinion on how this term should be defined and also concern that without a precise definition, student critiques and opinions were uninterpretable and uninformative. Faculty members and course directors also voiced concerns to the team about this outcome and often denied placing any emphasis on it in their development of course syllabi or conduct of classroom activities. (This seems unfortunate in light of the EOAWG's Final Report finding that a faculty member's relative emphasis on student attitudes was the strongest predictor ($r=.51$, $p<.001$) of students' subsequent rating of a course and teacher effectiveness.) Despite the discomfort and confusion concerning this outcome, the course contributions ranged from 4 to 7 with the majority of engineering core courses receiving ratings of 5.

2) ***What has the report told us about the contributions of specific core courses in the Engineering Division to the educational outcomes?***

As mentioned earlier, variation in the style, quantity and quality of the paragraphs supporting the ratings given it appears as though they were prepared by separate authors. If this is the case, drawing conclusions about the unique relative contributions of individual courses are very problematic. Also a persistent problem mentioned was that within many engineering core courses, raw data showed considerable variation between sections. This suggests that although the modal or normative section within a course may have been making adequate contributions, some sections with extraordinarily low ratings lowered the course average and increased the standard deviation.

Nonetheless, several general conclusions seem to emerge from a review of the ratings provided. The new integrative, interdisciplinary, problem-solving introduction to engineering principles and practices (Engineering 110), appears to be a tremendous success. The novel pedagogy and curricular design may provide a model for developments in other engineering core courses. (Unfortunately, the 100-Level Horizontal Assessment Team did not provide ratings of this experimental course, so external corroboration is not possible in this report.) Another persistent bright spot appeared to be those engineering core course substitutes taught just for technical majors; both Astro 320 and EE 231 were rated in the top half of the engineering courses. However, the core courses taught for non-technical majors (Astro 410 and EE 215) received below average assessment ratings (similar to the situation in Basic Sciences between Math 357 and Math 220).

3) ***What can be learned from the Engineering Division Assessment Report concerning the assessment process and practice in general?***

This divisional assessment reflects a significant effort by many dedicated individuals to provide accurate assessments of the contributions of each of the engineering division core courses. However, several external factors as well as internal processes appear to mitigate against confidence in these ratings. The overall average disagreement between divisional assessment ratings and those provided by the three Horizontal Assessment Team's ratings of course contributions is the highest of the four divisions (+.9) and the number and the number of individual "significant disagreements" (5) is higher than those encountered by the other three divisions combined (3).

As noted earlier, however, it appears as though course assessments were accomplished individually and relatively independently. Although some ratings appear to be mere assertions of adequacy others use convincing data and cogent argument to document the ratings they provide. The electrical and mechanical engineering courses appear to be good exemplars of this category. In contrast, three of the five significant disagreements with horizontal team assessments occurred in a single course (Civil Engineering 310). For each of the three outcomes the divisional rating was 2, and in one case 3, points higher on the rating scale than the horizontal teams assessment. The other two assessment discrepancies occurred in Engineering 310's contribution to students' ability to frame and resolve ill-defined problems and Engineering 410's contribution to the development of intellectual curiosity (another 3 point discrepancy). A review of the corresponding descriptive paragraphs for these ratings reveal a tendency to emphasize intention rather than evidence of execution effectiveness, and also a tendency to select a single supportive positive indicator or anecdote and ignore less supportive or contradictory data. Perhaps cadet participation on the horizontal teams was an effective remedy for these problems.

Chapter 7

100-Level Horizontal Assessment Team Report

CHAPTER SEVEN - 100 LEVEL HORIZONTAL ASSESSMENT TEAM REPORT

Ms. Dolores M. Karolick--34EDG
Maj Thomas Mabry--Department of Behavioral Sciences and Leadership
Capt Lisa Gossett--Office of the Registrar
C1C Cary Ellis--34TRW

OVERVIEW

The 100-level Core Course Horizontal Assessment Team was created by the Educational Assessments Working Group during its Phase II Charter to assess core course contributions across the curriculum at the 100 level in three areas; integrated knowledge, framing and resolving ill-defined problems, and intellectual curiosity. Contributions were made by this team in survey development, data analysis and report completion. Our intent was to review data collected with the expectation that 100 level core courses provide the basic knowledge, fundamental comprehension and initial college experience upon which the student's success in upper level coursework is established.

Twelve 100-level core courses, from each of the four divisions, were examined. They include, alphabetically by division and course:

Basic Sciences:	Chemistry 141 and 142, Computer Science 110, Math 141 and Math 142, and Physics 110
Engineering:	Engineering Mechanics 120
Humanities:	English 111, Foreign Language 132 and 142, History 101
Social Sciences:	Behavioral Science 110

Please note: Correlation of data collected was found to be challenging since the sample size for each division was not equal. Basic Sciences has six core courses, Humanities has four, and Engineering and Social Sciences have only one course a piece.

Survey Instruments

Four instruments were used to collect data. They are identified below in chronological order of usage:

1) 1995 Course Critiques.

   These critiques were accomplished by students at the end of both Spring and Fall semester in 1995. Five questions were selected from these critiques which best fit the three categories being assessed.

2) Survey of course practices.

This survey tool was accomplished by Course Directors for all core courses taught at USAFA. It asked questions concerning grading policies, types of class work, and instructional methodologies.

3) USAFA Education Outcomes Faculty Survey.

   This survey instrument was developed to survey faculty and course directors concerning their impression of how much emphasis is placed in a particular class on each educational outcome. This survey instrument was developed on a four point scale with 0=none, 1=light, 2=average, 3=heavy.

4) Specialized course survey.

   This tool was distributed only in core courses, near the end of the course, Spring semester 1996. The questions asked were significantly different from those asked on the standard course critique by emphasizing skills practiced for framing and resolving ill-defined problems, presentation of material for integrated knowledge, and intellectual curiosity. All but five questions have been used in prior studies with high reliability. The scale for this survey was: 1=strongly disagree, 2=disagree, 3=neutral, 4=agree, and 5=strongly agree.

The data is compiled and discussed in the following three sections, identified by the appropriate USAFA Educational Outcome. Each of the data collection tools will be discussed independently, with summary remarks at the end of each section.

METHODOLOGY

Collection of data through the survey tools described above provided interesting information, while difficult to compare. Since all tools have different response criteria, data from each tool have been looked at separately, with overall impressions identified as a summation.

As a starting point, the results of the *Survey of Course Practices* is first discussed. It provides an overview of each core course and policies surrounding the pedagogy incorporated. Following are overviews of the other three survey tools broken into the appropriate educational outcome category.

Survey of Course Practices

As a baseline for comprehending the course structure and presentation framework, course directors were requested to complete this survey which described course practices, policies and procedures. Questions asked were :

1. How is the final grade determined: % from final test, % from other tests, % from out of class work, % from collaboration?
2. What percent of questions on tests are multiple choice/short answer, essay, computational or performance in nature?
3. What percent of graded work is permitted to be reaccomplished for improvement?
4. For required assignments, do you allow students to select from an array of options for completion versus all students having the exact same assignment.

Following is a breakdown of responses by Division and Class for each 100-level core course reviewed.

Explanation of Columns:

Grade Composition:

Fnl%-Final Test Grade
GR%-Other Test Grades
Col%-Collaboration with other students
Out%-Out of Class Work graded

Question Composition:

MC - Multiple Choice/True-False
Essay - Essay
Comp - Computation
Perf - Performance
Redo - Reaccomplishment of work permitted
Opt - Students allowed to select from a list of topics/projects to accomplish

Course	*Grade Composition*				*Question Composition*					
	Fnl%	GR%	Col%	Out%	MC	Essay	Comp	Perf	Redo	Opt
Basic Sciences										
Chem141	30	45	0	9	0	20	80	0	0	0
Chem 142	27	41	0	14	3	18	79	0	0	0
CompSci 110	25	30	30	8	22	30	2	46	14	6
Math 141	26	47	15	19	38	21	41	0	0	0
Math 142	25	55	12	7	26	9	65	0	3	1
Engineering										
E Mech 120	28	41	5	21	18	6	64	12	15	0
Humanities										
Eng 111	0	0	75	0	0	75	0	25	8	60
FL 132	20	25	8	0	32	10	0	58	4	1
FL 142	23	29	17	24	11	70	0	19	20	22
Hist 101	25	20	30	0	26	69	0	5	3	62
Social Science										
Beh Sci 110	27	31	25	17	40	40	3	17	43	34

It is important to note that some course directors believe the grade composition, as described above, often over-lapped areas and therefore a 100% total was not attempted. For example, out of class work and collaboration often were involved with the same task and not easily broken apart.

INTEGRATED KNOWLEDGE

The first of the Educational Outcomes at the United States Air Force Academy encompasses a desire for a broad basic knowledge across the four divisions, with a depth of knowledge in an area of concentration. This team looked at how well each 100-level core course provided fundamental knowledge in the specific discipline as well as related/integrated that information into other disciplines. Following are results from three survey tools identified by their use chronologically.

1995 Course Critiques

Two questions from the Spring and Fall 1995 course critiques were selected to examine the question of integrated knowledge. The critiques utilized a response mechanism on a Likert scale from 1 to 6:

1=very poor 2=poor 3=fair 4=good 5=very good 6=excellent

Question 21, "Relevance and usefulness of course content was:" The mean score for this question for all 100-level core courses was 4.331.

Basic Sciences had scores ranging from 3.76 to 4.62, with an overall mean of 4.32

Engineering had a score of 3.99, just slightly below 4.0.

Humanities courses ranged in score from 4.07 to 4.5, with an overall mean of 4.36.

Social Science had a score of 4.61.

It appeared that the courses with the lower scores, still only slightly below the overall mean, were those in which students probably had less prior exposure to the subject matter, perhaps a computer programming or engineering class. Since the courses are introductory level, relevance to other subject areas may not be apparent to students at this early stage in their undergraduate work. These courses are intended to be a base line overview required of all students.

Relevance and usefulness of content seemed more obvious to students in subject areas where they have had extensive previous exposure; for example basic science or social science. When looking at the course practices, those classes which allowed flexibility of student selection of topics for course requirements and reaccomplishment of work scored highest.

Overall, students acknowledged the relevance of 100-level core course material to their academic careers and basic knowledge. Ratings from good to very good indicate a general appreciation for the contribution of the curriculum presented.

Question 22, "The instructor's effectiveness in facilitating my learning in the course was:" The mean score for this question for all 100-level core courses was 4.382.

Basic Sciences had scores ranging from 4.18 to 4.83 with an overall mean of 4.44

Engineering scored 4.14.

Humanities ranged from 3.98 to 4.59 with an overall mean of 4.29.

Social Science scored 4.63.

Social Science scored higher on overall score, but the highest scoring class came from the Basic Sciences at 4.83. Scores of this caliber indicate that students perceive faculty are involved with facilitating student learning in the classroom, are caring about student learning, and therefore, are working with students on their academic success.

USAFA Education Outcomes Faculty Survey

This survey instrument was developed to survey faculty and course directors concerning their impression of how much emphasis is placed in a particular class for each educational outcome. This survey instrument was developed on a four point scale:

0=none 1=light 2=average 3=heavy

The overall mean score was 2.795, indicating an extremely high emphasis on fundamental knowledge by faculty.

Basic Sciences ranged in score from 2.57 to 3.0, with an overall mean of 2.73.

Engineering scored 3.0.

Humanities ranged in score from 2.5 to 3.0, with an overall mean of 2.72.

Social Science scored 3.0.

Specialized Course Surveys

Questions eighteen through twenty-one specifically addressed integrated knowledge. The scale for this survey ranged from one to five:

1=strongly disagree 2=disagree 3=neutral 4=agree 5=strongly agree.

The questions asked were:

18. *The knowledge I have learned in this course has helped me better understand material in another core course.*
19. *This course routinely uses knowledge I have learned in another core course.*
20. *I have noticed an attempt in this core course to integrate material with other core courses.*
21. *This core course contributes to the creation of an integrated body of fundamental knowledge.*

The overall mean, all courses and all questions, was 3.22, indicating a tendency to just slightly above a neutral position. Standard deviation remained fairly consistent across the data. Following is a compilation of results by division.

Division	Question 18	Question 19	Question 20	Question 21	Overall
Basic Science	3.07	3.22	3.24	3.52	3.26
Engineering	3.38	3.52	3.42	3.55	3.47

Humanities	3.00	2.83	2.86	3.62	3.08
Social Science	3.05	2.83	2.90	3.57	3.09
TOTALS	3.13	3.10	3.11	3.57	3.22

Students identified through this survey that they observed no real evidence, either positively or negatively, that information obtained in 100-level core courses helped them with material in other courses or that there was an attempt to relate information between courses. They agreed that information contained in 100-level core courses in all divisions contributed to a larger body of integrated, fundamental knowledge.

Summation

All three survey instruments indicate an agreement by both faculty and students that 100-level core courses contribute to a body of fundamental knowledge. Students do not always acknowledge usefulness of specific information, however, or integration of information to other subject areas. This may be due to the level and inexperience of the students, since they are at a freshman college level, to judge from a larger perspective.

It is clear that faculty have high expectations about fully meeting USAFA Educational Outcome #1. The integration of both the fundamental knowledge and integration of that knowledge between disciplines may be falling short of intent. A greater emphasis may need to be placed in each course concerning how information presented relates to other fields.

FRAMING AND RESOLVING ILL-DEFINED PROBLEMS

The second USAFA Educational Outcome, the ability to frame and resolve ill-defined problems, is described as the ability to deal with problems that are ambiguous, interactive and ever-changing. Framing means constructing a working model, and revising it based on feedback. Resolving means that an ill-defined problem is never solved for good; rather it is solved again and again (re-solved) as the problem is framed again and again; and each successive solution is more refined (resolution.)

1995 Course Critiques

Two questions from the Spring and Fall 1995 course critiques, question 16 and 36, were selected to examine the question of framing and resolving ill-defined problems. The critiques utilized a response mechanism on a Likert scale from 1 to 6:

1=very poor 2=poor 3=fair 4=good 5=very good 6=excellent

16. Intellectual challenge and encouragement of independent thought were:

The mean score for this question for all 100-level core courses was 4.58.

Basic Sciences had scores ranging from 4.39 to 4.93, with an overall mean of 4.63.

Engineering had a score of 4.35.

Humanities had scores ranging from 4.49 to 4.55, with an overall mean score of 4.51.

Social Science had a score of 4.79.

Extremes did not exist on these scores or means. Students agreed that core courses at the 100-level did a good to very good job of challenging them intellectually and encouraging them toward independent thought.

36. This course improved my ability to deal with problems that don't have an "approved solution." *The overall mean score for this question at the 100-level was 4.94.*

Basic Sciences had scores ranging from 4.56 to 5.31, with an overall mean of 4.95.

Engineering had a score of 4.67.

Humanities had scores ranging from 4.82 to 4.96, with an overall mean of 4.89.

Social Science had a score of 5.36.

Students perceive this area as "very good," in all divisions. Our committee found this to be amazing at the 100-level, since in many cases, courses at this level are thought to contain only "well-defined" problems, memorization of facts, and very little ill-defined direction. Receiving a score at this level indicates, that students believe, the institution has taken seriously the need for future leaders and graduates of the Academy to have this skill and to begin challenging student abilities immediately.

USAFA Education Outcomes Faculty Survey

Faculty and course directors completed this survey to provide their impressions of how much emphasis is placed in 100-level core courses on, in this case, framing and resolving ill-defined problems. This survey instrument was developed on a four point scale:

0=none 1=light 2=average 3=heavy

The overall mean score was 1.76, indicating an average to light emphasis by faculty on ill-defined problems at the 100-level.

Basic Sciences ranged in score from 1.07 to 2.40, with an overall mean of 1.61.

Engineering scored 1.25.

Humanities ranged in score from 1.0 to 2.09, with an overall mean of 1.53.

Social Science scored 2.67.

This survey indicated a great disparity between divisions concerning emphasis on framing and resolving ill-defined problems at the 100-level. Social Science expressed a high necessity and inclusion of this type of work in their curriculum, while Engineering indicated a need for defined problems and more absolute information. This diversity of opinion between disciplines prepares students for problem solving at both ends of the spectrum, providing intellectual opportunities which otherwise might not be explored.

Specialized Course Surveys

The majority of questions asked on this survey directly related to critical thinking skills and framing and resolving ill-defined problems. Questions seven through eleven and thirteen

through sixteen specifically addressed this issue. The scale for this survey ranged from one to five:

1=strongly disagree 2=disagree 3=neutral 4=agree 5=strongly agree.

The questions asked were:

7. In this course I learned useful strategies for approaching complex questions in a variety of reasonable ways.

8. In this course I frequently found myself actively engaged in thinking about difficult questions for which we still need to find answers.

9. In this course I improved my ability to evaluate new information and analyze the central ideas of this subject area.

10. In this course I improved my ability to give sound reasons for my beliefs and opinions regarding issued in this subject area.

11. As a result of taking this course I find I am more fair-minded.

13. As a result of taking this course my thinking is more focused and systematic, at least in this subject area.

14. The instructor encouraged thoughtful exploration of the central ideas and relationships in the course content.

15. The way the instructor conducted this course illustrated how to think in reasonable, objective, and fair-minded ways.

16. The assignments (tests, readings, projects, papers, classroom activities) in this course frequently engaged me in complex thinking.

	Question #									
Division	7	8	9	10	11	13	14	15	16	Overall
Basic Science	3.64	3.51	3.72	3.34	3.05	3.59	3.77	3.74	3.76	3.57
Engineering	3.73	3.43	3.76	3.34	2.96	3.66	3.95	3.78	3.82	3.60
Humanities	3.21	3.21	3.63	3.45	3.28	3.43	3.69	3.58	3.37	3.43
Social Science	3.54	3.66	3.69	3.73	3.40	3.47	3.56	3.55	3.58	3.58
TOTALS	3.53	3.45	3.70	3.47	3.17	3.54	3.74	3.66	3.63	3.54

Students indicated that they were neutral about being more fair-minded as a result of the course. On the other hand, they tended to agree with the fact that instructors encouraged their thoughtful exploration of central ideas and relationships in content and that they improved their ability to evaluate new information and analyze central ideas. Six of the questions were rated between neutral and agree, indicating a tendency toward learning useful strategies in problem solving, active engagement in thinking about difficult questions, providing sound reasons for opinions, and thinking in a more focused/systematic way.

Summation

Students and faculty seemed to disagree in the area of framing and resolving ill-defined problems and the skills being introduced and practiced which promote this ability. Students found courses to be thought provoking, challenging them with complex questions and requiring them to organize and substantiate data, and to form opinions based upon that data.

Faculty, on the other hand, did not believe that a tremendous emphasis was being placed on requiring the skill of framing and resolving ill-defined problems at the 100-level. Their focus tended toward other Educational Outcomes, such as fundamental knowledge. Perhaps the difference in opinions has to do with level of expertise in the subject matter. Students found challenge in questions that faculty assumed were fundamental. Sophistication within the discipline may be the differentiating factor when it comes to this observation.

INTELLECTUAL CURIOSITY

The final area to be explored by this team was the sixth USAFA Educational Outcome, intellectual curiosity. This outcome is defined as going beyond possessing knowledge and the abilities to put it to use. An intellectually curious student is one who is *inclined* to explore; to develop an *attitude* that predisposes him or her to lifelong learning. Less data was collected on this topic than either of the other two outcomes. Following is a summation of the three survey tools incorporated.

1995 Course Critiques

One question from the Spring and Fall 1995 course critiques was selected to examine the question of intellectual curiosity. The critiques utilized a response mechanism on a Likert scale from 1 to 6:

1=very poor 2=poor 3=fair 4=good 5=very good 6=excellent

39. There are a number of things in this general subject area I'd like to read more about. The mean score for this question for all 100-level core courses was 4.73.

Basic Sciences had scores ranging from 4.10 to 5.15, with an overall mean of 4.58.

Engineering had a score of 4.06.

Humanities had scores ranging from 4.42 to 5.39, with an overall mean of 4.92.

Social Science had a score of 5.32.

This question brought out the greatest diversity of opinion between divisions. While scores at their very lowest were still at a 4 or "good" rating, this area also posted the highest scores of approaching "excellent." In even the worst cases, we observed there was interest sparked in a majority of students to pursue more information in general subject areas. At the undergraduate freshman level this is inspiring.

USAFA Education Outcomes Faculty Survey

Faculty and course directors were asked to complete this survey to provide their impressions of how much emphasis is placed in 100-level core courses intellectual curiosity. This survey instrument was developed on a four point scale:

0=none 1=light 2=average 3=heavy

The overall mean score was 1.81, indicating an average to light emphasis by faculty on developing intellectual curiosity at the 100-level among students.

Basic Sciences ranged in score from 1.47 to 2.10, with an overall mean of 1.69.

Engineering scored 1.5.

Humanities ranged in score from 2.0 to 2.27, with an overall mean of 2.09.

Social Science scored 2.83.

Other than Social Science where a heavy emphasis is placed on intellectual curiosity, the other three divisions indicated an average to light emphasis. This is exemplified by the overall 100-level mean of 1.81, indicating below average to light emphasis. This attitude is consistent with responses given earlier where faculty placed their greatest emphasis on integrated knowledge. While intellectual curiosity is considered a nice side effect, it is apparently not a key concern of the 100-level faculty.

Specialized Course Surveys

Only one question on this survey dealt with intellectual curiosity. The scale for this survey ranged from one to five:

1=strongly disagree 2=disagree 3=neutral 4=agree 5=strongly agree.

Questions asked was:

12. As a result of taking this course my interest and curiosity about the issues and questions in this subject area has grown. The overall mean score for all 100-level core courses was 3.40.

Basic Sciences had scores ranging from 3.10 to 3.50, with an overall mean of 3.27.

Engineering scored 3.17.

Humanities had scores ranging from 3.12 to 4.30, with an overall mean of 3.58.

Social Science scored 3.62.

Students seemed less enthusiastic about issues and questions in specific subject areas during this survey than in the 1995 course survey. The expression seemed to be in a positive direction, but not overwhelming, which coincides very closely with what the faculty told us during the same period.

Summation

It appears from the data collected that while intellectual curiosity is an important issue for some faculty, it is not considered to be as critical at the 100-level as other educational

outcomes. This was reflected in the attitudes of the 1996 student critiques in which attitudes were much less enthusiastic than those expressed by the 1995 student critiques. It would have been very interesting to see what the faculty emphasized in 1995 courses in reference to this area. Perhaps the data would have correlated directly with student and faculty attitudes during that time frame.

CONCLUSION

This study revealed that, as may be expected, students and faculty often disagree about what is happening in the classroom. Faculty placed a great emphasis on fundamental knowledge. Students observed that same emphasis, however, without attempting to integrate that knowledge into other disciplines. Faculty are less concerned at approaching ill-defined problems and intellectual curiosity at the 100-level. Students, however, observed these courses as intellectually challenging, requiring them to think and stretch in new directions. This difference may be a function of "novice" versus "expert" perspectives.

It is difficult to determine whether this study has identified essential similarities and/or differences between divisions and courses with reference to USAFA Education Outcomes. It may be a greater contribution to identify concerns from this study's methodology and suggested short falls in our approach. It is clear that several considerations should be accounted for in the future. Some of those issues are:

1. Survey tools should attempt to use the consistent Likert scale. Direct statistical analysis of data would be more meaningful, descriptive, and valuable.
2. Same population sizes might also be attempted. Having different number of courses represented in each division, and extremely diverse numbers of students represented, feasibly paints an inaccurate conclusion about the data collected. While the total number of subjects for the study was impressive, representative student populations varied widely between divisions.
3. Survey tools need to be examined for clarity, ease of use, and universal definition. Faculty surveys were considered complex and confusing by some participants.
4. The Educational Outcomes Assessment Working Group attempted to pull together information utilizing existing data, which may or may not have always been appropriate. Having data collected from different times, (Spring 1995 and 1996), may skew the results.

Further studies need to be accomplished with defined parameters to clearly answer the original questions. It may be stated that the apparent direction for all divisions at the 100-level is positive for all three Educational Outcomes reviewed. None of the data pointed in a negative direction, or indicated a lack in attempt to include each educational outcome in each core course. The mere fact that these outcomes are being examined and recognized by both faculty and staff is encouraging and exciting since identification of these outcomes is new at the Air Force Academy.

100-Level Core Course Comparison

Previous sections of this report have combined data within divisions to report comparisons. This section reports comparison data between individual 100-level core courses by each of the three

educational outcomes explored. Ratings are based on a scale from one to seven. These ratings are defined as:

1 = unsatisfactory level of contribution
2 = poor contribution
3 = slightly below average contribution
4 = satisfactory level of contribution
5 = slightly above average contribution
6 = excellent contribution
7 = best possible contribution

Results from the three survey instruments; 1995 course critiques, specialized course survey, and faculty survey have been combined to obtain the following mathematical ratings. Scores are displayed from lowest to highest for all three categories.

Integrated Knowledge

Computer Science 110	- 4.90	Foreign Language 132	- 5.35
English 111	- 4.90	History 101	- 5.40
Foreign Language 142	- 5.00	Engineering Mech 120	- 5.45
Physics 110	- 5.10	Behavioral Science 110	- 5.50
Chemistry 142	- 5.15	Math 142	- 5.60
Math 141	- 5.35	Chemistry 141	- 5.70

Framing and Resolving Ill-defined Problems

History 101	- 4.50	Foreign Language 132	- 4.90
Chemistry 141	- 4.55	Math 141	- 5.00
Engineering Mech 120	- 4.75	English 111	- 5.25
Chemistry 142	- 4.75	Computer Science 110	- 5.45
Math 142	- 4.80	Physics 110	- 5.80
Foreign Language 142	- 4.85	Behavioral Science 110	- 5.90

Intellectual Curiosity

Engineering Mech 120	- 4.40	English 111	- 5.00
Math 142	- 4.50	Foreign Language 142	- 5.35
Chemistry 142	- 4.75	Foreign Language 132	- 5.45
Chemistry 141	- 4.80	Physics 110	- 5.50
Computer Science 110	- 4.80	History 101	- 5.65
Math 141	- 4.85	Behavioral Science	- 6.00

Summary

It is important to note that this is a baseline study, identifying beginning points for all courses. With this in mind, it is critical to comprehend that in each category, courses were rated between satisfactory and above average and that two of the three survey instruments were accomplished by students. While we would like to see all courses rated at a "7" in all three areas, it is encouraging to confirm that courses are at this level today, with an expectation to strive toward excellence in the future.

100-Level Commentary and Critique

1) ***What has this report revealed about the three educational outcomes for 100 (i.e., freshmen) level core courses?***

This team's average ratings of 12 freshman level courses revealed that integrated knowledge did slightly better (5.3) than did framing and resolving ill-defined problems (5.0) or developing (or retaining) cadets' intellectual curiosity (5.1). However, the distribution of scores shows much greater range of course contributions for intellectual curiosity (1.6) and framing and resolving (1.4) than integrated knowledge (.8). Several interesting issues were revealed by the team's discussions of each of these outcomes.

While students appeared willing to accept that content from different courses did inter-relate they were somewhat unsure as to how. Many of these potential connections, in contrast, appeared obvious to many faculty members. The team's comments concerning the disproportionate distribution of courses from the four academic majors also begs the question of just what we as an institution mean by integrated knowledge. Recent curricular changes (both of the sole representatives from Social Sciences and Engineering academic divisions (Psychology and Mechanical Engineering) were moved from 100 to 200 level courses) might suggest that only courses from the Basic Sciences and Humanities divisions are a part of freshmen integrated knowledge and that Social Science and Engineering courses are opportunities to apply these basics. One consequence of this perspective is the lost opportunities to use courses from Social Sciences and Engineering to integrate concepts and principles from the Humanities and Basic Sciences. One example of this is the new problem-centered experimental Engineering 110Z course (unfortunately not directly evaluated by this team). The development of such a course for the social sciences warrants careful consideration. The General Psychology course, which applied scientific principles and practices to issues of broad significance addressed by Humanities, also provided an opportunity to demonstrate connections between the academic divisions.

To a much greater extent than faculty, students recognized and appreciated the contributions of many freshmen core courses to their ability to frame and resolve ill-defined problems. This surprised the team and is just the opposite of the situation with integrated knowledge where faculty saw the connections but students didn't. Perhaps this is the other side of the same coin. Because relations both among and within freshmen courses are much less clear to students than to faculty members, students are likely to interpret new (but technically "well-defined") problems as being "ill-defined." It also seems reasonable that many of the skills students use with such problems do prepare them to later deal with actual rather than only apparently "ill-defined" problems.

As noted above, the intellectual curiosity outcome showed the greatest diversity in ratings of course contributions. Ironically, this may be because contributions to intellectual curiosity is much easier to compare than integrated knowledge which may be fundamentally incomparable (how many equations equate to a sonnet, etc.). Overall intellectual curiosity seemed to be important, but faculty struggled with the lack of a common, agreed upon definition (perhaps an "ill-defined " problem in its own right).

2) ***What has the report told us about the contributions of specific 100 level core courses to the educational outcomes?***

Individual course ratings are revealed in an attachment at the end of the report. These were used along with some of the comments and explanations in the body of the report which dealt primarily with combined divisional inputs. Math 141 (5.7) and Chemistry (5.6) distinguished themselves by their contributions to integrated knowledge. The sole course representatives also did relatively well on this outcome (Behavioral Sciences 110 (5.5) and Engineering Mechanics 120 (5.45)). Computer Science 110 and English 111 received 4.9 ratings, but this difference does not appear to be a problem in absolute terms (4.0 = Average) nor relative to the 5.26 overall average rating for this team.

As far as the framing and resolving ill-defined problems outcomes, Behavioral Sciences 110 (5.9) and Physics 110 (5.8) distinguished themselves from the other freshmen courses. In contrast, History 101 (4.50) and Chemistry (4.55) were rated significantly lower than the average. It is striking how different courses which might be considered to have a somewhat similar content and even pedagogy could receive such diverse ratings. Actual contributions to cadet development of these skills might have much less to do with divisional or disciplinary perspective than with pedagogical or epistemological perspectives (viz., how much educational ambiguity is appropriate). Behavioral Science 110 also received the top score for contributions to intellectual curiosity (6.0) followed by History 101 (5.65) and Physics (5.5). Engineering Mechanics 110s (4.4) and Mathematics 142s (4.5) were rated at the bottom. This may be an opportunity for improvement since most cadets express relatively high interest in the various engineering programs upon their arrival at the Academy. A more positive first impression, one that showed how hard work could lead to deep satisfaction, might lead to more sustained interest by a greater number of cadets.

Simply averaging scores across these three outcomes yields a global measure of overall contributions. Ten of the twelve courses examined fell within .25 of the "above average" rating of 5.0. Behavioral Sciences 110 and Physics 110 distinguished themselves on the positive side with overall average contributions of 5.80 and 5.47 respectively, tops among the freshmen core.

3) ***What can be learned from the 100 Level Horizontal Assessment Report concerning the assessment process and practice in general?***

This report appears to reflect an objective and systematic approach. The thorough discussion of what raw data was used was also very helpful in understanding the decisional process. The decision to rate course contributions to the nearest .05 created the appearance of much greater precision; but, at least in the case of integrated knowledge, may have reduced the overall range. The reflections at the end of the report are particularly commendable and reflect an authentic awareness of the assessment process. The fact that no members of the team were from the Basic Sciences or Engineering departments was a disadvantage. However, no one from the Engineering Division volunteered to serve on any of the Horizontal Assessment Teams, and in general, membership of Horizontal Teams tended to avoid individuals from departments with one (or more) core courses being considered by the particular team. This is not to suggest that the team was unqualified to assess the contributions of the technical core courses to the outcomes but

to concede that such assessments are likely to have relied more heavily upon student data than did assessments by the divisional teams. It is also likely that the team's assessment of technical courses' contributions to students' intellectual curiosity and ability to frame and resolve ill-defined problems is likely to be relatively more valid than their assessment of each courses' contributions to students' knowledge.

Several minor problems with these ratings were apparent: 1) no courses were rated either flat average or below average, 2) grouping by academic division seemed to obscure more about individual course contributions than it revealed and 3) although mentioned in the overall summary, non-random and non representative samples across divisions distorted the meaning of the raw averages presented in the body of the report.

Chapter 8

200-Level Horizontal Assessment Team Report

CHAPTER EIGHT - 200 LEVEL HORIZONTAL ASSESSMENT TEAM REPORT

Capt Sandra Eisenhut--Department of Behavioral Sciences and Leadership
Capt Dean Steele--Department of Philosophy and Fine Arts
Capt Crystal M. Jonas--Office of the Registrar

Purpose of 200-Level Core Courses

Due to USAFA's core course requirements, all students would have been exposed to every academic division, and most academic departments by the end of their sophomore year. In general, course directors (CDs) of 200-level core courses indicated the purpose of the 200-level core is to provide a continuation of the 100-level experience designed to "build the basics" and give the students fundamental knowledge in various subject areas. Since all but one of the 200-level core courses have prerequisites (most have two), 200-level courses are more integrative in nature.

Some faculty members feel the role of the upper-level courses is to provide specialized knowledge in a specific domain; and therefore, the lower level courses should provide some integration across disciplines. Regardless of whether "integration" is supposed to happen in lower level courses or upper level courses (or both), course directors and instructors of 200-level core courses felt their relative emphasis was on the educational outcome related to knowledge versus developing either intellectual curiosity or critical thinking skills (with the exception of English 211). This perspective is evident in the course directors' and instructors' responses to the educational outcomes relative emphasis questionnaire. For the purposes of this report, data were only extracted for the three outcomes of knowledge, critical thinking and intellectual curiosity (see Figure 1). On a scale of 0 (no coverage) to 3 (heavy coverage), the overall 200-level rating by CDs and instructors on building an integrated, fundamental knowledge was 2.5 (as compared to 2.7 for 100-level courses and approximately 2.26 for 300 and 400-level courses). Course directors' and instructors' ratings of their emphasis on the outcomes related to framing and resolving and intellectual curiosity reflected an opposite pattern of emphasis: in freshman core courses, the emphasis on framing and resolving skills was 1.6, and 1.9 for 200-level courses (approximately 2.4 for 300/400 levels). The educational outcome of developing students who are intellectually curious received a "light to average" emphasis by 100-level courses (1.79), but was emphasized more in the 200-level courses (2.26--between average and heavy). The 200-level emphasis on intellectual curiosity was rated higher (though probably not statistically significant) than the 1.89 emphasis in the 300 and 400 level courses. (See Attachment 8, p. 31).

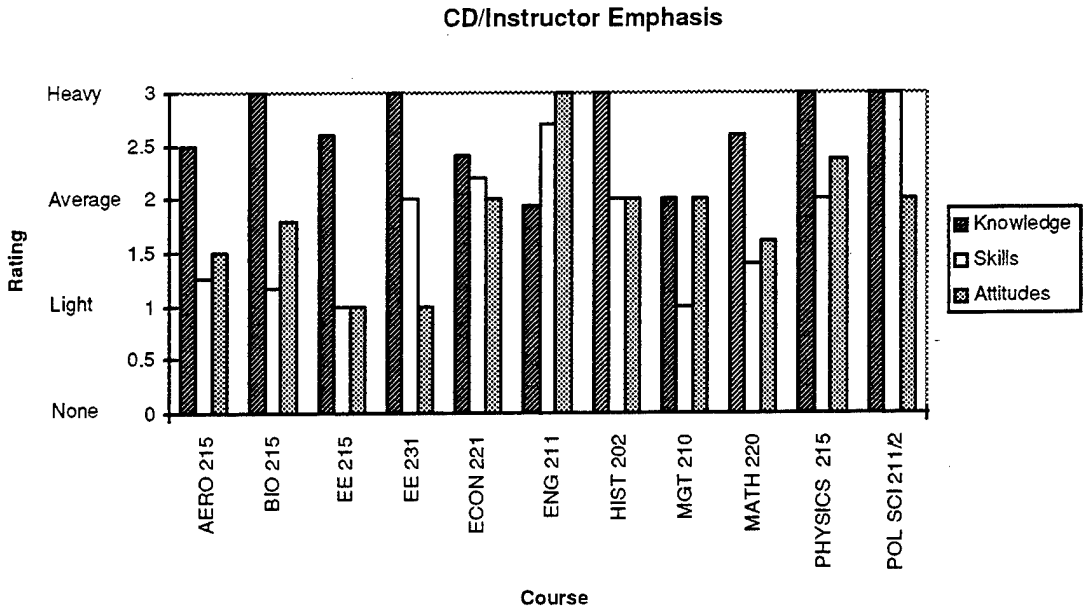


Figure 1. 200-level Course Director and Instructor ratings of their courses

Figure 1 emphasisizes the educational outcomes related to knowledge, framing and resolving skills, and intellectual curiosity (attitudes).

Note that the ratings from course directors and instructors on the relative emphasis for the educational outcomes are somewhat "unstable" for two reasons. First, the number of responses for each course varied greatly; some courses had only one input, while others had as many as eight. Second, the definition of the outcome may have been ambiguous, especially for the outcome related to "knowledge." It is unclear from the data whether CDs and instructors were rating their emphasis to provide "fundamental knowledge" or "integrated knowledge," or both. Despite the instability in the CD/instructor responses, the data are valuable in terms of understanding what the goals of the 200-level courses are in comparison to the students' perceptions of the courses' contributions.

Assessment

General Approach. Each of the 200-level core courses was rated by a committee of three faculty members (each from a different academic department) using the standard levels of 1-7 (with possible increments of .5) proposed by the Faculty Educational Outcomes Assessment Working Group. While there was an abundance of data, the most useful data were those obtained from the Course Assessment survey given during April-May 1996 to students enrolled in core courses.

The Core Assessment data confirmed trends indicated in Course Critique data from the previous two years relating to questions regarding intellectual challenge (Question 16), relevant content (Question 21), amount learned (Question 22), the amount of problems without approved solutions (Question 36), and reading more about a subject outside of class

(Question 39). The Course Critique data are depicted in Appendix A, and other than noting their convergence with the Core Assessment student data, were not addressed further.

Data collected from CDs on *what* their practices were, by percent, for a course (e.g., number of GRs, collaborative effort, reaccomplishment of work, etc.) were less useful because they did not address "how" these practices were implemented. The trends in the practices data did not seem to reflect correlations with student perceptions of the educational outcomes, suggesting accomplishments of the educational outcomes may be more linked to *how* instructors implement classroom practices and less to *what* or *how much*. (Note that the 100-, 200-, and 300/400-level Assessment Teams did not interview CDs. CD interviews were conducted by Divisional Assessment Teams.)

The Course Assessment Survey. Data from the Course Assessment survey instrument reflected a different number of student responses, as shown in Table 1. The least number of responses for a course was 34, which on a practical level, is sufficiently large. Therefore, no statistical tests were done to mediate the affects of the different sample sizes.

Table 1
Number of Responses on the Course Assessment

Course	Responses
Aero 215	34
Bio 215	215
EE 215	336
EE 231	82
Econ 221	295
Engl 211	445
Hist 202	331
Mgt 210	284
Math 220	129
Physics 215	389
Pol Sci 211	175
Pol Sci 212	466

Questions 7-21 from the Course Assessment were associated with one of the three educational outcomes addressed. Specifically, each question is listed below in Table 2 and cross-referenced to the outcome it addressed:

Table 2
Course Assessment Questions Categorized by Educational Outcome

Question	Knowledge
18	The knowledge I have learned in this course has helped me better

	understand material in another core course.
19	This course routinely uses knowledge I have learned in another core course.
20	I have noticed an attempt in this core course to integrate material with other core courses.
21	This course contributes to the creation of an integrated body of fundamental knowledge.
Question	**Skills (Framing and Resolving)**
7	In this course I learned useful strategies for approaching complex questions in a variety of reasonable ways.
9	In this course I improved my ability to evaluate new information and analyze the central ideas of this subject area.
10	In this course I improved my ability to give sound reasons for my beliefs and opinions regarding issues in this subject area.
11	As a result of taking this course I find I am more fair-minded.
13	As a result of taking this course my thinking is more focused and systematic, at least in this subject area.
15	The way the instructor conducted this course illustrated how to think in reasonable, objective, and fair-minded ways.
16	The assignments (tests, readings, projects, papers, classroom activities) in this course frequently engaged me in complex thinking.
17	This course lends itself to ill-defined problems (issues without a single approved solution)
Question	**Attitudes (Intellectual Curiosity)**
8	In this course I frequently found myself actively engaged in thinking about difficult questions for which we still need to find answers.
12	As a result of taking this course my interest and curiosity about the issues and questions in this subject area have grown.
14	The instructor encouraged thoughtful exploration of the central ideas and relationships in the course content.

Each of the questions were rated by students on the following scale:

Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree
1	2	3	4	5

Responses to the survey questions for each outcome are grouped and charted by course in Figures 2, 3 & 4. The mean 200-level response for questions relating to each educational outcome is indicated by a gray shaded area. The mean course response across all questions for each outcome is indicated by a black horizontal bar.

It is interesting to note that the averaged student ratings (Knowledge: 3.14; Skills, 3.43; and Attitudes: 3.46) reflect benchmarks which are only slightly above "neutral" according to the above scale. In other words, the average student ratings indicated that the students only slightly agreed that their 200-level core courses, as a whole, made contributions to their

learning a fundamental, integrated knowledge base, developed their skills at framing and resolving ill-defined problems, or developed their intellectual curiosity!

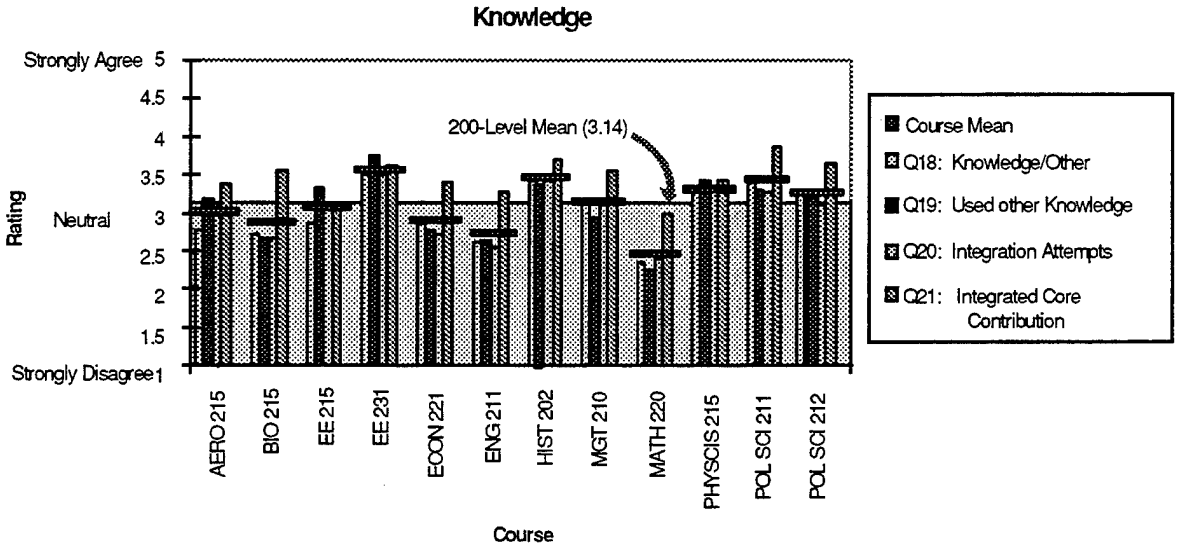


Figure 2. Course Assessment responses relating to the KNOWLEDGE educational outcome by course.

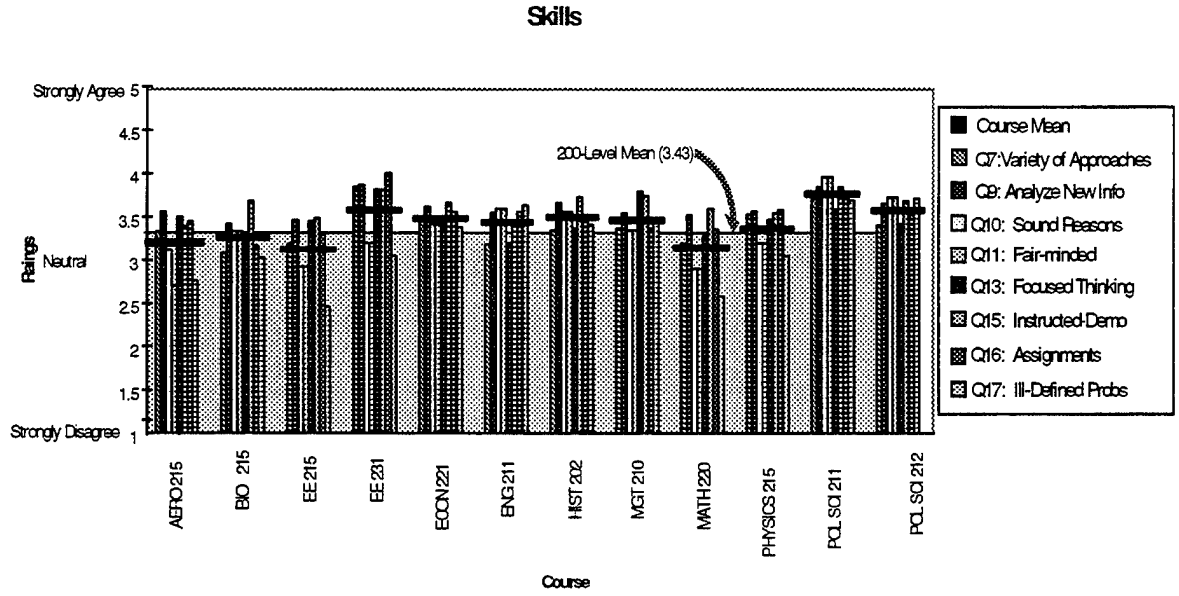


Figure 3. Course Assessment responses relating to the FRAMING AND RESOLVING SKILLS educational outcome by course.

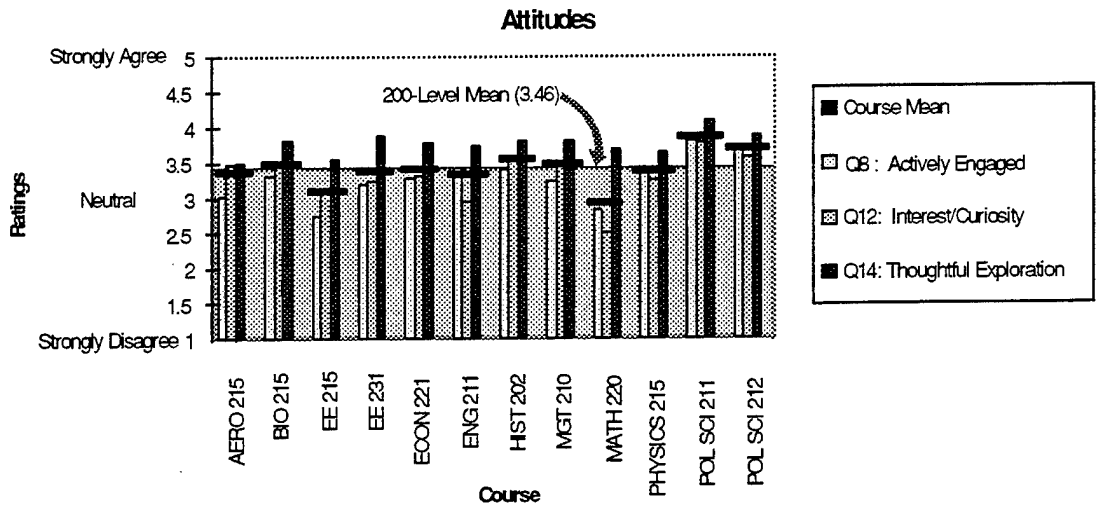


Figure 4. Course Assessment responses relating to the INTELLECTUAL CURIOSITY (ATTITUDES) educational outcome by course.

The three committee members rated each course on how the student perceived the course as contributing to the educational outcomes relating to knowledge, framing and resolving skills and intellectual curiosity (developing the students' attitude). Each member compiled the data from the different available sources, primarily using the Course Assessment data. The ratings for each of the committee members did not vary by more than 2 levels for each course, indicating relative consensus across independent raters. The ratings were averaged, and rounded to the nearest ½ scale increment. It is important to recognize these ratings as a reflection of the students' perceptions of the course contributions to each of the educational outcomes. The raters did not make assumptions about what the "expected" level of emphasis for each of the courses should be, e.g., no assumption was made that each course was designed to contribute equally to the knowledge, skills and attitudes of the students.

Ratings

AERO 215, Fundamentals of Aeronautics

Prerequisites: EngrMech 120, Physics 110, Comp Sci 110

Fundamental Integrated Knowledge: 4.5

Framing and Resolving Skills: 4.5

Intellectual Curiosity/Attitudes: 4.5

CD/instructors rated their emphasis for developing knowledge as "2.5 (heavy emphasis)," and emphasized framing and resolving skills and developing intellectual curiosity comparatively less (1.25 and 1.5, respectively—which reflect "less than an average" amount). The students, however, perceived the relative contribution of the course to all three about equally, at a level slightly below the 200-level averages.

In Aeronautics 215, the final grade for students in this course is comprised of 13% collaborative effort (the average collaboration extent across all core courses); 52% of the final grade was based on computations (25% was the average)--perhaps not a noteworthy amount, given this is an engineering course; 43% of the final grade in the course was based on work done by students outside of class (more than one standard deviation above the overall core course mean of 21.8%, (sd= 15.0); and the extent of options for students was only 5% (the core course average was 12%).

Overall, this course is rated as between average to slightly above average in terms of contributing to the educational outcomes of knowledge, skills and attitudes.

BIO 215, Introductory Biology with Lab (2 hours)

Prerequisites: Chem 111/112, 131 or 151

Fundamental Integrated Knowledge: 4

Framing and Resolving Skills: 4

Intellectual Curiosity/Attitudes: 4.5

The Course Director and Instructors in Biology 215 indicate that the primary emphasis of the course is to build a fundamental, integrated knowledge, rating this emphasis at the top of the scale with a "3." Relatively less emphasis is placed on developing framing and resolving skills (1.17-light to average) and moderate emphasis on developing students' intellectual curiosity (1.80-slightly below average).

Students rated Biology 215 as slightly below the 200-level average for developing their integrated, fundamental knowledge base, and also slightly below the average for contributing to their ability to frame and resolve ill-defined problems (not more than one standard deviation from the mean for either outcome). However, the course was rated by students as slightly above the average for developing their intellectual curiosity (students: 3.53; 200-level average 3.46--again, not more than a standard deviation from the average). This corresponds to Course Critique data for the last two years that indicates students generally agree that this course inspires them to read more about this subject area outside of the course requirements.

Approximately 27% of the work in Biology 215 involved collaborative effort; however the extent that students could exercise their own options was only 10% (the overall core average was 12.1%), and the course did not allow reaccomplishment of work to bring up a grade.

Overall, the ratings for this course indicate a satisfactory (average) contribution to the educational outcomes related to an integrated, fundamental knowledge, framing and resolving skills, and student intellectual curiosity.

ECON 221, Principles of Microeconomics

Prerequisites: Math 141

Fundamental Integrated Knowledge: 3.5

Framing and Resolving Skills: 5.5

Intellectual Curiosity/Attitudes: 4.5

The relative emphasis of the three educational outcomes for Economics 221 was about equal--with an average emphasis placed on each (knowledge: 2.4; skills: 2.20; attitudes: 2.00).

The students rated the contributions of knowledge and attitudes close to the average of all 200-level courses (a neutral rating), but rated the contribution of the course to framing and resolving ill-defined problems as slightly higher (although not a standard deviation higher). The students' overall grade in the course was comprised of 50% multiple choice questions (more than one standard deviation higher than all core courses); 44% of the overall grade was determined by essay-type questions (none of the overall grade depended on making computations with an approved solution) and 6% of the overall grade was determined by cadet performance (e.g., participation, building, making, etc.). The percentage of out-of-class work was approximately 17.5%; collaboration was allowed on only 2% of the course, and there were not opportunities to have "options" on assignments or to reaccomplish work to bring up a grade.

Overall, the strongest contribution Microeconomics makes to the three outcomes is developing the ability of students to frame and resolve ill-defined problems (rated as slightly above average). The contribution to student intellectual curiosity is satisfactory, but the student perception of how the course contributes to their integrated, fundamental knowledge is between satisfactory and slightly below satisfactory.

EE 215, Electrical Signals and Systems

Prerequisites: Math 142

Fundamental Integrated Knowledge: 5

Framing and Resolving Skills: 4.5

Intellectual Curiosity/Attitudes: 4

The CD/Instructor emphasis was on contributing to the student's knowledge (2.6/3.0: heavy) over framing and resolving abilities (1.0: light) and intellectual curiosity (1.0: light).

The students were neutral about this course's contribution to their ability to frame and resolve ill-defined problems and their intellectual curiosity. Students did rate the course as slightly above average in contributing to their fundamental knowledge.

Electrical Engineering 215 was the second highest course (second to EE 231) in the 200-level core in terms of computations with approved solutions making up the student's grade (73%). This is more than one standard deviation greater than the overall core average. The course also used 0% multiple choice (below one standard deviation from the overall core) and only 22% essays without computations (the overall core average is 35.6%, sd=22.8%). None of the grade was comprised of work performed outside of the classroom. Collaborative efforts were 8% of the grade (overall core average = 13.6%, sd= 17.9%) and approximately 3% of the overall grade input was allowed to be reaccomplished. None of the courses allowed the student options to choose from assignments or projects.

Overall, Electrical Engineering 215 was assessed to provide satisfactory to slightly above average contributions to the educational outcomes related to knowledge, skills and attitudes.

EE 231, Electrical Circuits and Systems I (core substitute for EE 215)

Prerequisites: Math 142

Fundamental Integrated Knowledge: 6

Framing and Resolving Skills: 5.5

Intellectual Curiosity/Attitudes: 4.5

The CD of Electrical Engineering 231 placed the greatest amount of emphasis on providing an integrated, fundamental knowledge (3.0/3.0: heavy); an average amount of emphasis on developing students' skills to frame and resolve ill-defined problems (2.0/3.0: average); and the least amount of emphasis on developing intellectual curiosity (1.0/3.0: light).

In contrast to EE 215, students "slightly agree" that EE 231 contributed to their integrated fundamental knowledge (3.6; 200-level mean = 3.14, sd=.39). This is the only course in the 200-level core to be rated by students as greater than one standard deviation above the 200-level mean for contributing to their fundamental knowledge. It is also the 200-level core course whose overall grade uses the largest amount of computations with approved solutions (78% - more than a standard deviation above the mean for all core courses). Interestingly, despite the high percentage of computations with approved solutions in the course, the students "slightly agree" (3.6/5.0) the course develops their ability to frame and resolve ill-defined problems. The students agreed that the assignments in this course engaged them in complex thinking--the highest rating on question 16 of the Course Assessment survey of all 200-level core courses. Sixteen percent of the course was comprised of assignments/work accomplished outside of the class (below the overall core average of 21.8%, sd=15%), no collaborative effort was allowed, and no portion of the graded work was allowed to be reaccomplished for a higher grade. Only 4% of the course permitted the student options.

Among the 200-level core courses, Electrical Engineering 231 was the top contributor (assessed as "excellent" by the assessment team) to providing an integrated, fundamental knowledge. (In absolute terms, the students said they slightly agreed that it made a contribution.) Students rated this course as the second highest (second to DFPS 211) contributor of all 200-level core courses which developed their ability to frame and resolve ill-defined problems. In general, using the standard EOAWG rating scale, this course is slightly above average to excellent in its contributions to the educational outcomes of knowledge, skills and attitudes.

ENG 211, Masterpieces of Literature

Prerequisites: English 111

Fundamental Integrated Knowledge: 3

Framing and Resolving Skills: 5

Intellectual Curiosity/Attitudes: 4

English 211 was the only 200-level core course in which the CD and instructors indicated the relative emphasis (1.93/3.0: very light) on building an integrated, fundamental knowledge was less than the emphasis on the framing and resolving skills (2.71/3.0: average-heavy), and developing intellectual curiosity (3.0/3.0: heavy).

The student ratings seem to validate the instructors' emphasis, being highest for framing and resolving and intellectual curiosity (3.47/5 and 3.38/5, respectively) and lowest for knowledge (2.77/5.0). The student rating for knowledge was ranked last of the 200-level courses, and was close to, but not exceeding, one standard deviation below the 200-level course mean.

Not surprisingly, the course required no computations, and 81% of the course was based on essays (more than one standard deviation above the overall core course mean of 35.6%). It is likely that contribution of English 211 to developing framing and resolving skills, rated slightly above the average for the other 200-level courses, is attributable to the 45% of the course which allowed students to have options with respect to assignments (this percentage is second only to History 202). A less expected finding was that no graded portion of the course was allowed to be reaccomplished for a higher grade.

Overall, English 211 was assessed as providing a slightly below average contribution to the students' integrated, fundamental knowledge, but satisfactory to slightly above average contributions to intellectual curiosity and framing and resolving skills, respectively. It may be true that the student's perception of what they are gaining in their knowledge base from English 211 (given the heavy emphasis of the Academy toward Engineering/Basic Sciences) is not fully appreciated by the students. This presents some intriguing questions for exploration at both the course level (e.g., how do we help the student understand the importance of this subject area?) and the institution level (should every course "do everything"? or do we need to assess the student perception of a soldier-scholar-citizen and the well-roundedness this implies?).

HIST 202, Introduction to Military History

Prerequisites: History 101

Fundamental Integrated Knowledge: 5.5

Framing and Resolving Skills: 5.5

Intellectual Curiosity/Attitudes: 5

The CD's judgment of History 202's emphasis on knowledge was 3.0/3.0 (heavy); 2.0/3.0 (average) for developing framing and resolving skills as well as developing student intellectual curiosity.

The student ratings on the Course Assessment placed this course above average for all 200-level courses for its contributions to all three outcomes. The mean rating for questions relating to contributions to an integrated, fundamental knowledge was 3.5/5.0 almost one standard deviation above the 200-level mean (M=3.14; sd=.39), making it the highest contributor by this metric.

History 202 is most remarkable for the level of options given to students with respect to assignments and projects--an estimated 58% of the course permitted flexibility and choice for the students. This is more than two standard deviations above the overall core average (M=12.1%, sd=20.8%), placing it at the top of the list for 200-level courses and a close second to English 211 (60%) in all core courses. Three percent of the course grade was made up of assignments students could reaccomplish to bring up their grades; none of the graded assignments were collaborative. More than half of the course grade was essay in nature; 43% was based on multiple choice, and the remaining 5% was based on student performance (participation, contributions, etc.).

Overall, the course was rated as making above average contributions to the three educational outcomes addressed.

MGT 210, Introduction to Management

Prerequisites: None

Fundamental Integrated Knowledge: 4.5

Framing and Resolving Skills: 5

Intellectual Curiosity/Attitudes: 5

The CD's estimated degree of emphasis on providing integrated, fundamental knowledge and developing student intellectual curiosity was average (2.0/3.0), with relatively lower emphasis on developing framing and resolving skills (1.0/3.0: light).

Student responses to the Course Assessment survey indicate they felt "neutral" to "slightly agreed" that Management 210 contributed to these educational outcomes. Of the 200-level courses, Management 210 based the largest portion (27%) of the students' grade on performance (contributions, participation, etc.); 50% of the final grade was based on multiple choice questions; and the remaining 23% of the course grade was based on essay questions. This course also leads all 200-level core courses in the extent of collaborative effort (35%)--which is greater than one standard deviation higher than the overall core mean of 13.6% (sd=17.9%). Management 210 is the number one core course overall for in the amount of out-of-class work (50%, M=21.8%; sd=15.0%). While none of the assignments are allowed to be reaccomplished to bring up grades, 19% of the course grade is comprised of assignments where the students have options/choice, and 35% of the course allows collaborative effort (more than one standard deviation above the overall core mean, M=13.6,sd=17.9%).

Using the EOAWG standardized scale, Management 210 is rated as making slightly above average contributions to knowledge, skills and attitudes of students.

MATH 220, Probability and Statistics

Prerequisites: Math 142 or department permission

Fundamental Integrated Knowledge: 3

Framing and Resolving Skills: 4

Intellectual Curiosity/Attitudes: 3.5

The CD and instructors of Math 220 judged the emphasis of their course to be heaviest for providing an integrated, fundamental knowledge (2.6/3.0: heavy), and a relatively lighter emphasis on developing student intellectual curiosity and framing and resolving ill-defined problem skills (1.6/3.0 and 1.4/3.0, respectively). Student ratings on the Course Assessment survey indicated that Math 220 was one standard deviation below the 200-level average for contributing to an integrated knowledge (2.5/5.0; M=3.14, sd=.39), and one standard deviation below the average rating for developing student intellectual curiosity (3.01/5.0; M=3.46, sd=.35), and close to one standard deviation below the mean for developing skills (3.17/5.0; M=3.43,sd=.30). Students did slightly disagree that the course lent itself to ill-defined problems (2.59/5.0); interestingly, the relative order of contributions toward the three outcomes were: thinking skills (3.17/5.0), attitudes (3.01), followed by knowledge(2.50/5.0).

The student's grade in Math 220 is proportioned as 39% computations with approved solutions, 40% multiple choice inputs, 19% essay, and 2% student performance (participation, contributions, etc.). Students are not allowed to reaccomplish any work to bring their grades up and have no "options" in selecting projects or assignments. The CD estimated 6% of the course is collaborative and 12% of the work was out-of-class.

The committee's overall ratings suggest that further inquiry may be needed, and opportunities for improvement should be given careful consideration.

PHYSICS 215, General Physics II (2 hours)

Prerequisites: Physics 110, Math 142

Fundamental Integrated Knowledge: 5

Framing and Resolving Skills: 4

Intellectual Curiosity/Attitudes: 3.5

For Physics 215, the CD/instructor order of emphasis on the three educational outcomes addressed were: first, providing an integrated, fundamental knowledge (3.0/3.0: heavy); second, developing intellectual curiosity (2.30/3.0: average to heavy); and third, developing framing and resolving skills (2.0/3.0: average).

Student ratings on the Course Assessment indicated their perceptions of this course's contribution to an integrated, fundamental knowledge as slightly above the average for the 200-level mean (3.37/5.0), and the contributions toward framing and resolving skills and intellectual curiosity approximately at the mean (3.39/5.0 and 3.44, respectively).

Five percent of the course was collaborative, 5% of the course permitted students to reaccomplish assignments, 5 % of the course was categorized as "out of class" work, and 5% of the course allowed students to have options for projects or assignments. The portion of the final grade comprised of computations with approved solutions was 44%; 35% was multiple choice; 18% was essay; and 3% of the student's final grade was determined by performance (contributions, participation, etc.).

In general, Physics was rated as providing a slightly above average contribution to fundamental knowledge, an average contribution to developing framing and resolving skills, and slightly below average to average contribution to developing intellectual curiosity.

DFPS 211	***DFPS 212***
Politics and American Government	***International Politics and U.S. National Security***
Prerequisites: None	**Prerequisites: Pol Sci 211**
Fundamental Integrated Knowledge: 5.5	Fundamental Knowledge: 5.5
Framing and Resolving Skills: 6	Framing and Resolving: 5.5
Intellectual Curiosity/Attitudes: 6	Intellectual Curiosity/Attitudes: 5.5

Political Sciences 211 and 212 can be considered as one, year-long course (per the Political Science Core Course Director). Thus, the two courses will be discussed together. These two courses are logically sequenced and integrated well with each other. The CD/instructor ratings of the educational outcomes indicate a heavy emphasis on knowledge and framing and resolving skills (both 3.0/3.0) and an average emphasis on developing intellectual curiosity (2.0/3.0).

The students' responses to the Course Assessment indicated Political Sciences 211 and 212 contributions to their integrated, fundamental knowledge, skills and attitudes were above average for all 200-level courses. Political Science 211 was rated one standard deviation above the other courses for developing framing and resolving skills (3.79; M=3.43, sd=.30).

Both courses have slightly more than half the grade input from essays, approximately 44% of the grade input from multiple choice questions, and the remaining 5-6% based on performance (contributions, participation, etc.). Both courses use 35% work outside of the class (the overall core course mean is 21.8%), and approximately 10% of each course is collaborative. Neither course gives student "options" for assignments or allows reaccomplishment of graded activities.

The convergence of the Course Critique data with the Course Assessment ratings provide compelling evidence that Political Sciences 211 and 212 are making Above Average to Excellent contributions to student knowledge, skills and attitudes and are rated among the top of the 200-level core courses.

Conclusions

The ratings provided by the committee were based on the standard rating scales used by the Educational Outcomes Assessment Working Group. The data collected did not necessarily allow exact matches to the benchmarks provided by the rating scales, but every attempt was made to make meaningful interpretations of what was available during this limited time frame.

Most conclusions were drawn from trends identified in the student responses collected with the Course Assessment.

When trying to make sense of these data, it is important to note that the average ratings given on the Course Assessment indicated that students were only slightly above neutral (toward agreeing) for all of the courses' contributions toward knowledge, skills and attitudes.

Given the limitations of the data, the limited sensitivity of the measurement tool, and the purpose of this assessment, it doesn't make sense to rank order the 200-level courses by contributions to the educational outcomes under study. Alternately, in line with the goal of continuous improvement, it makes sense to look at courses that stand out as making noticeable positive or negative impacts on the educational outcomes to identify practices which help or hinder our purpose so we can affect positive changes. With this intention in mind, Table 3 is provided to summarize the 200-level ratings. The average committee ratings across all 200-level core courses for each of the outcomes were:

Knowledge:	4.58 (sd=1.02)
Skills:	4.92 (sd= 0.61)
Attitudes:	4.63 (sd= 0.68)

Table 3
Summary of 200-level Assessment Ratings
* indicates ratings 1 standard deviation above the mean rating
# indicates ratings 1 standard deviation below the mean rating

Course	Knowledge Mean=4.59 sd=1.02	Skills Mean=4.92 sd=0.61	Attitudes Mean=4.63 sd=0.68
Aero 215	4.5	4.5	4.5
Bio 215	4	#4	4.5
EE 215	5	4.5	4
EE 231	*6	*5.5	4.5
Econ 221	#3.5	5	4.5
Engl 211	3	5	4
Hist 202	5.5	*5.5	5
Mgt 210	4.5	5	5
Math 220	#3	#4	#3.5
Physics 215	5	4.5	4.5
Pol Sci 211	5.5	* 6	* 6
Pol Sci 212	5.5	*5.5	*5.5

200-Level Commentary and Critique

1) ***What has this report revealed about the three educational outcomes at the 200 level of core courses?***

As students transition from freshmen core courses to the sophomore or 200 level core courses, faculty members' relative emphasis on the three educational outcomes shifts. Although still high in both relative and absolute terms, emphasis on integrated knowledge is slightly lower than in freshmen courses. Emphasis on students' ability to frame and resolve ill-defined problems increases significantly, but is still less than it is in 3 and 400 level courses. Emphasis on intellectual curiosity is at about the same level as emphasis on framing and resolving ill-defined problems. However, unlike the emphasis on framing and resolving which continues to increase, emphasis on intellectual curiosity actually appears to peak at the sophomore level. Perhaps this is because most students select their academic major during the sophomore year and there is an additional parochial advantage to presenting the discipline in the best possible light until that time. Another even more obvious explanation is that during the junior and senior years, core courses must compete with course demands in students' chosen academic major.

This assessment team's absolute rating averages across the 12 core courses considered in their report show a similar pattern: overall contributions to integrated knowledge and intellectual curiosity are slightly above average (4.58 and 4.63 respectively). Rated contributions to the critical thinking outcome (framing and resolving was slightly higher (4.92). Unlike the 100 level team, this team seemed to have little trouble differentiating relative contributions to integrated knowledge (sd = 1.02) while it was much less willing to disperse ratings of course contributions to either framing and resolving ability (sd = .61) or intellectual curiosity (sd = .68).

2) ***What has the report told us about the contributions of specific 200 level core courses to the educational outcomes?***

This team's willingness to use a wider range of ratings to describe core course contributions made these differences very salient. For example, the integrated knowledge ratings of 6.0 garnered by Electrical Engineering 231 and the 5.5s awarded to History 202 and Political Science 211 and 212 reflect relatively strong and consistent contributions. In contrast, the 3.0s awarded to Math 220 and English 211 as well as the 3.5 given to Econ 221, reflect the team's consensus that there was room for improvement in this area. It is interesting to note that three of the four academic divisions are represented among the top and bottom contribution categories with respect to integrated knowledge.

Course contributions to students' ability to frame and resolve ill-defined problems identified several positive exemplars: Political Science 211 (6.0) and Political Science 212, Economics 221, History 202 and Electrical Engineering 231 (all at 5.5). Although most of this group of courses are from the realm of social science, the inclusion of an engineering course is noteworthy. No courses were rated below average for their contributions in this area but the three lowest ratings of 4.0 all went to Basic Science courses: Math 220, Physics 215 and Biology 215.

Once again Political Science 211 was identified as the course contributing the most (6.0) to students' intellectual curiosity, followed closely by Political Science 212 (5.5) and then by History 202 and Management 210 at 5.0. Math 220 and Physics 215 both received slightly negative assessments (3.5) on this outcome and Electrical Engineering 215 and English 211 received neutral (4.0) ratings for their contributions to this outcome.

Averaging contributions across the three outcomes provided a general measure of outcome contributions. Both Political Science courses (211 and 212) received very high average contribution ratings (5.83 and 5.50 respectively) followed closely by Electrical Engineering 231 at 5.33. Electrical Engineering 231 is a substitute for the general Electrical Engineering core course (215) which received an only slightly above average rating of 4.50. On the down side, only one course was rated negatively overall: Math 220 received an average contribution rating of 3.5. The other two Basic Science courses didn't fare much better, both receiving only neutral scores of 4.00.

In selecting courses to be assessed, a deliberate attempt was made to be inclusive rather than exclusive. Although Electrical Engineering 231 is a course designed for only a few students (viz., those majoring in electrical engineering or computer science), it was included in the overall analysis because it serves as a substitute for Electrical Engineering 215. Were electrical engineering to be excluded from the analysis, ratings for the other science and engineering courses would be relatively low. Whether this pattern is attributable to the students' perceptions, faculty performance, or assessment team bias or methodology is unclear. The fact that Electrical Engineering 231 received such high ratings, however, suggests that the team was willing to recognize excellence in a technical course when sufficient evidence existed.

3) ***What can be learned from the 200 Level Horizontal Assessment Team Report concerning the assessment process and practice in general?***

This is another good example of a very systematic and thorough approach to assessment. The use of graphical representations of the data considered helps communicate the distribution of scores fairly clearly. It also is laudable that concerns about unequal and/or non representative sample sizes are reported as appropriate caveats. In each of the course assessment discussions, unique curricular practices based on course directors' responses to structural questions are identified and considered as they might relate to the other data and evidence of overall contributions. However, the authors openly admit at the onset that the results of this analysis do not provide any apparent pattern. Apparently, the structural distinctions captured by the curricular survey are not the most important influences on course contributions to the educational outcomes. The thorough and explicit discussion of the independent individual rating process and consensus building is another notable strength of this report. The relatively high level of initial consensus among raters (no difference greater than 1.5) reflects a common understanding of the rating scale as well as similar perspectives on the relative value of the various sources of data.

Chapter 9

Upper-Level Horizontal Assessment Team Report

CHAPTER NINE - UPPER LEVEL HORIZONTAL ASSESSMENT TEAM REPORT

Captain Andrew P. Armacost--Department of Management
Captain David E. Bell--Department of Physics
Major Robert Steigerwald--Department of Computer Science
Captain Peter Pastor--Department of Chemistry
2nd Lieutenant Justin Hansen--USAFA Class of 1996, Department of Behavioral Sciences and Leadership

1. Introduction

Prior to the Spring Semester, 1996, assessment efforts at the U.S. Air Force Academy focused on individual courses, courses grouped by department, and courses grouped by academic division. Divisional assessment groups were formed to address the contribution to seven primary educational outcomes by each of the four divisions: Social Sciences, Humanities, Engineering, and Basic Sciences. Complementing the divisional groups during the Spring 1996 semester were newly formed horizontal assessment working groups. One group explored the 4-degree experience, the second grouped explored the 3-degree experience, and the third group explored the experience of cadets in the upper level core courses. The findings of this third group are the subject of this report.

The initial analysis of the upper level core courses, comprising any core course designated in the 300's or the 400's, began with the assembly of existing data (i.e., end-of-course critiques and course syllabi) and the collection of new data (i.e., course design questionnaires for instructors and student questionnaires that supplemented the end-of-course critiques). Given the large amount of data, summaries of the course ratings for each of the 300- and 400-level courses were generated, from which composite ratings of the courses were derived.

Revisions of the Assessment Working Group's charter at the beginning of the semester narrowed the focus of this assessment effort upon three educational outcomes:

> *Outcome #1:* Officers who possess breadth of integrated, fundamental knowledge in basic sciences, engineering, humanities, and social sciences (a.k.a., Fundamental Knowledge).
> *Outcome #2:* Officers who can frame and resolve ill-defined problems (a.k.a., Ill-defined Problems).
> *Outcome #6:* Officers who are intellectually curious (a.k.a., Intellectual Curiosity).

The purpose of this report is to develop overall ratings for each of the 300- and 400-level core courses with respect to the three educational outcomes of interest. This paper will focus on the methodology used to transform data from several sources into single ratings for each course relative to the three outcomes. General findings and conclusions are presented in Section 5 and assess the individual courses and the general trends of the entire 3/400 level.

2. Data Collection Methodology

To assess core course contributions to the DF educational outcomes, various tools and data sets were employed. Existing data from past assessment surveys were supplemented with questionnaires administered this semester. This section presents the instruments used for data collection and divides them into three categories: data sets that were used for this assessment, data sets that were not used for this assessment, and data that would have been useful if they were available.

2.1. Data/Instruments Used in 3/400 Horizontal Assessment

Three primary data collection instruments were used to develop the composite measures: Student-submitted end-of-course critiques, student- and instructor-submitted supplemental course assessment surveys, and course director- and instructor-submitted course design coverage surveys. Each is described below.

2.1.1. End-of-Course Critiques

Results from two previous semesters of USAFA end-of-course critiques were tailored to only include data relevant to the three DF educational outcomes of interest. Questions 16 (Intellectual challenge), 21 (Relevancy of Course Content), 22 (Amount Learned), 36 (Problems without an approved solution), and 39 (Read more about subject) relate to one or more of the three outcomes. For each course, an average value was calculated for each question. These averages are presented in Section 3.

2.1.2. Supplemental Course Assessment Survey

The questions from the end-of-course critiques were somewhat limited in relevance to the educational outcomes. As a result, a Supplemental Course Assessment Survey was administered and generated additional data. This survey includes standard Facione questions to help measure the development of critical thinking skills and intellectual curiosity (Questions 7-16). Additional questions include a single question on how the course lends itself to ill-defined problems (Question 17) and a set of questions relating to Fundamental Knowledge and the integration of such knowledge (Questions 18-21). Again, course averages were calculated for each question. These averages are presented in Section 3.

2.1.3. Course Design Coverage Questionnaire

The two previous instruments relied heavily on student responses. The Course Design Coverage Questionnaire was given to core course directors and instructors who evaluated their course's coverage of the seven DF educational outcomes. The coverage scale ratings (from 0 *[none]* - 3 *[high emphasis]*) are the primary measures of interest. The remaining categories in this instrument were subject to such widespread misinterpretation that the results were meaningless. For the coverage scale, a single course average was given for each educational outcome. These averages are included in Section 3.

2.1.4. Course Syllabi

Course syllabi from the Spring 1996 semester were collected for each of the 300- and 400-level courses. These provided a qualitative look at the design and approach of each course.

2.2. Data/Instruments Not Used in 3/400 Horizontal Assessment

Several data collection instruments that may have been used by other horizontal or divisional assessment teams were not used in the 3/400 assessment. A description of each and the rationale for not including them our composite ratings are provided in this section.

2.2.1. College Basic Academic Subjects Examination

This standardized test was administered to 4th class cadets and 1st class cadets in 1990. The differences in scores highlight both the successes and failures of the Academy to increase knowledge in a variety subject areas. Since the report shows changes in proficiency in *subject areas* over the course of a four-year education, this instrument is useful to identify performance grouped by subject area, but that does not provide insight into the upper-level courses' contributions.

2.2.2. Grading Strategy Form

Each core course director developed a profile of how points were allocated in their courses. This addressed the numbers of tests and Graded Reviews and it also included the manner in which students were graded (i.e., essay, multiple choice, computational questions). Correlating these measures with the educational outcomes requires considerable speculation. For example, one might infer that if a course relied completely on multiple choice questions, then the course focuses heavily on Fundamental Knowledge when, in fact, the questions might *not* test Fundamental Knowledge. The uncertainty of these grading methods and their relationship to the outcomes was too great to include this instrument in the 3/400 composite scores.

2.3. Other methods

Finally, since the process of assessment is a continual endeavor, the following additions are suggested for follow-on work: create additional questions to help fill the voids from previously used instruments; continue to use existing instruments so that time-series trends can be used to track increasing and decreasing emphasis on the educational outcomes; and convene small focus groups for each of the horizontal assessment working groups so that a subjective component can be incorporated into the large amounts of survey data. A note of caution must be sounded: before pressing ahead with new data collection, careful analysis of existing data should be accomplished. One example is to compare student versus instructor versus course director responses on the Supplemental Course Assessment Survey (Section 2.1.2 and Appendix D). This provides a good comparison of how well the faculty's intentions are actually satisfying the outcomes from the students' perspectives. A final recommendation is to collect data from "employers" of USAFA graduates and the graduates themselves. For the graduates, capturing their feedback within four years of graduation is important, since they would recall the curriculum and would be able to assess its impact upon their Air Force careers.

3. Summary of Raw Data

Each instrument discussed in the previous section generated data relating to the three educational outcomes of interest: Fundamental Knowledge, Framing and Resolving Ill-Defined problems, and Intellectual Curiosity. This section will present, in tabular form, the raw data for each outcome. The approach will be to describe which data were used from each of the three primary instruments: Course Design Coverage Ratings (done by Course Directors and Instructors), End-

of-Course critiques (done by students), and Supplemental Course Assessment (done by students and instructors). Note that specific data or questions may relate to more than a single educational outcome.

3.1. Data for Outcome #1: Fundamental Knowledge

The following items are taken from the Course Design Coverage Ratings (referred to in Table 1 as "A"):

- ❑ Outcome #1 - Rating

The following items are taken from the End-of-Course Critiques (referred to in Table 1 as "B"):

- ❑ *Question #16* - Intellectually Challenging
- ❑ *Question #21* - Relevance of Course Material
- ❑ *Question #22* - Amount Learned

Finally, the following items were taken from the Supplemental Course Survey (referred to in Table 1 as "C"):

- ❑ *Question #9* - In this course I improved my ability to evaluate new information and analyze the central ideas in this subject area.
- ❑ Question #10 - In this course I improved my ability to give sound reasons for my beliefs and opinions regarding issues in this subject area.
- ❑ *Question #13* - As a result of taking this course my thinking is more focused and systematic, at least in this subject area.
- ❑ *Question #18* - The knowledge I have learned in this course has helped me better understand material in another course.
- ❑ *Question #19* - This course routinely uses knowledge I have learned in another core course.
- ❑ *Question #20* - I have noticed an attempt in this core course to integrate material with other core courses.
- ❑ *Question #21* - This core course contributes to the creation of an integrated body of fundamental knowledge.

The raw averages for each upper level courses, along with the 3/400 and DF averages, are shown in Table 1.

Division	Course	A1	B16	B21	B22	C9	C10	C13	C18	C19	C20	C21
Engineering	Astro 320*	2.80	4.40	4.00	4.40	3.48	3.18	3.38	2.61	3.02	2.72	3.22
	Civ Eng 310	2.00	3.50	3.25	3.70	3.01	2.92	2.83	2.05	2.26	2.27	2.59
	Engr 310	2.80	4.15	3.60	3.95	3.53	3.10	3.48	2.85	2.96	2.93	3.25
	Astro 410	2.86	4.00	3.50	3.90	3.51	3.29	3.36	2.72	3.18	2.88	3.37
	Engr 410	1.62	3.50	3.15	3.00	3.30	3.29	3.06	2.66	3.22	3.30	3.15
Humanities	Engl 311	2.00	4.30	4.20	3.80	3.48	3.69	3.13	2.76	3.07	2.84	3.32
	Phil 310	2.14	4.55	4.30	4.10	3.67	3.66	3.36	2.70	2.54	2.58	3.42
Social Sci	Beh Sci 310	2.31	4.80	5.20	4.90	3.63	3.73	3.47	3.32	3.18	3.20	3.78
	Econ 310		4.70	4.65	4.60	3.69	3.55	3.66	3.40	3.44	3.44	3.69
	Law 320	3.00	4.70	5.05	5.00	3.83	3.87	3.70	3.04	2.85	2.89	3.91
	Law 420	2.50	5.00	5.55	5.25	3.99	3.95	3.75	3.28	3.45	3.27	3.89
	3/400 Avg	2.40	4.42	4.27	4.33	3.56	3.47	3.40	2.87	3.06	2.97	3.43
	USAFA Avg	2.53				3.59	3.40	3.41	2.99	3.05	3.00	3.45

Table 1: Summary of Averages from Questions
Relating to Educational Outcome #1
(* denotes core substitute courses)

3.2. Data for Outcome #2: Ill-defined Problems

The format used in the previous section will be used with respect to the second educational outcome: Framing and Resolving Ill-defined problems. From the Course Design Coverage Survey, the single item of interest is:

- Outcome #2: Rating

The following question relating to ill-defined problems were taken from the End-of-Course critiques:

- *Question #36:* Problems without an approved solution

Finally, the following questions were taken from the Supplemental Course Survey:

- *Question #7:* In this course I learned useful strategies for approaching complex questions in a variety of reasonable ways.
- *Question #8:* In this course I frequently found myself actively engaged in thinking about difficult questions for which we still need to find answers.
- *Question #9:* In this course I improved my ability to evaluate new information and analyze the central ideas of this subject area.
- *Question #10:* In this course I improved my ability to give sound reasons for my beliefs and opinions regarding issues in this subject area.
- *Question #13:* As a result of taking this course my thinking is more focused and systematic, at least in this subject area.
- *Question #14:* The instructor encouraged thoughtful exploration of the central ideas and relationships in the course content.

- *Question #15:* The way the instructor conducted this course illustrated how to think in reasonable, objective, and fair-minded ways.
- *Question #16:* The assignments (tests, readings, projects, papers, classroom activities) in this course frequently engaged me in complex thinking.
- *Question #17:* This course lends itself to ill-defined problems (issues without a single, approved solution).

The raw averages for each upper level courses, along with the 3/400 and DF averages, are shown in Table 2.

Division	Course	A2	B36	C7	C8	C9	C10	C13	C14	C15	C16	C17
Engineering	Astro 320*	2.20	4.90	3.36	3.16	3.48	3.18	3.38	3.61	3.49	3.53	3.25
	Civ Eng 310	2.80	4.75	2.88	2.93	3.01	2.92	2.83	3.29	3.05	2.76	3.62
	Engr 310	1.00	4.60	3.32	3.08	3.53	3.10	3.48	3.72	3.72	3.49	2.77
	Astro 410	1.14	4.40	3.26	3.23	3.51	3.29	3.36	3.80	3.74	3.38	2.94
	Engr 410	3.00	4.75	3.43	3.46	3.30	3.29	3.06	3.37	3.42	3.16	4.03
Humanities	Engl 311	2.20	4.80	3.16	3.37	3.48	3.69	3.13	3.53	3.49	3.45	3.47
	Phil 310	2.86	5.30	3.71	3.98	3.67	3.66	3.36	3.89	3.74	3.45	4.14
Social Sci	Beh Sci 310	2.56	5.75	3.80	3.69	3.63	3.73	3.50	4.09	4.11	3.48	3.94
	Econ 310		5.30	3.55	3.79	3.69	3.55	3.66	3.98	3.90	3.64	3.67
	Law 320	2.67	5.45	3.65	3.69	3.83	3.87	3.70	3.77	3.72	3.52	3.76
	Law 420	2.67	5.95	3.88	3.75	3.99	3.95	3.75	4.07	4.06	3.74	3.86
	3/400 Avg	2.31	5.12	3.47	3.48	3.56	3.47	3.40	3.75	3.69	3.45	3.57
	USAFA Avg	1.96		3.42	3.41	3.59	3.40	3.41	3.72	3.62	3.51	3.38

Table 2: Summary of Averages from Questions
Relating to Educational Outcome #2
(* denotes core substitute courses)

3.3. Data for Outcome #6: Intellectual Curiosity

The approach used in the previous two sections will be used with respect to Educational Outcome 6: Developing Intellectual Curiosity. From the Course Design Coverage Survey, the single item of interest is:

- Outcome #6: Rating

The question relating to intellectual curiosity was taken from the End-of-Course critiques:

- *Question #39:* Would you read more on this subject?

Finally, the following question was taken from the Supplemental Course Survey:

- *Question #12:* As a result of taking this course my interest and curiosity about the issues and questions in this subject area has grown.

The course averages for these three items are summarized in Table 3.

Division	Course	A6	B39	C12
Engineering	Astro 320*	2.20	5.30	3.61
	Civ Eng 310	1.50	3.75	2.40
	Engr 310	1.40	4.15	3.19
	Astro 410	2.00	4.50	3.48
	Engr 410	2.00	3.70	2.59
Humanities	Engl 311	2.20	4.30	2.91
	Phil 310	2.14	4.75	3.23
Social Sci	Beh Sci 310	2.13	5.65	3.67
	Econ 310		5.30	3.62
	Law 320	2.00	5.60	3.78
	Law 420	2.33	5.80	3.90
	3/400 Avg	1.99	4.82	3.32
	USAFA Avg	1.93		3.29

Table 3: Summary of Averages from Questions Relating to Educational Outcome #6 (* denotes core substitute courses)

4. Composite Ratings for 3/400 Level Courses

Given the raw data, composite ratings were developed according to the seven-point scale agreed upon by the assessment teams. These ratings combine data from the three primary instruments discussed in Sections 2 and 3. This section will first present the methodology for arriving at the composite ratings and will then present the ratings themselves. Finally, general findings and conclusions regarding each core course's contribution to the three educational outcomes are presented.

4.1 Methodology

The composite scores were derived according to the general assessment scale. However, the data collected from the three primary instruments described in the previous section were insufficient to be able to rate the courses in strict accordance with the definitions. The particular difficulty arose when trying to assess consistency across all sections of the courses. Thus, the following methodology, using the scale from one through seven, was used to rate the each course *with respect to other courses in the 300- and 400-levels*. Conclusions that can be drawn from the ratings will be discussed in Section 4.3

Sections 3.1 - 3.3 list the questions that were used from each instrument for each educational outcome. First, as a measure of each course's performance on each instrument, ratings were averaged to give a *single measure for each instrument* (as shown in Table 4). Next, since each instrument used a different rating scale, these averages were *normalized* between the values of

zero and one (as shown in Table 5). Finally, the three normalized averages *were averaged to give a single measure* to represent that course's performance on each educational outcome (as shown in Table 6).

Division	Course	Outcome #1 Instrument A	Outcome #1 Instrument B	Outcome #1 Instrument C	Outcome #2 Instrument A	Outcome #2 Instrument B	Outcome #2 Instrument C	Outcome #6 Instrument A	Outcome #6 Instrument B	Outcome #6 Instrument C
Engineering	Astro 320*	2.80	4.27	3.09	2.20	4.90	3.38	2.20	5.30	3.61
	Civ Eng 310	2.00	3.48	2.56	2.80	4.75	3.03	1.50	3.75	2.40
	Engr 310	2.80	3.90	3.16	1.00	4.60	3.36	1.40	4.15	3.19
	Astro 410	2.86	3.80	3.19	1.14	4.40	3.39	2.00	4.50	3.48
	Engr 410	1.62	3.22	3.14	3.00	4.75	3.39	2.00	3.70	2.59
Humanities	Engl 311	2.00	4.10	3.18	2.20	4.80	3.42	2.20	4.30	2.91
	Phil 310	2.14	4.32	3.13	2.86	5.30	3.73	2.14	4.75	3.23
Social Sci	Beh Sci 310	2.31	4.97	3.47	2.56	5.75	3.77	2.13	5.65	3.67
	Econ 310		4.65	3.55		5.30	3.71		5.30	3.62
	Law 320	3.00	4.92	3.44	2.67	5.45	3.72	2.00	5.60	3.78
	Law 420	2.50	5.27	3.65	2.67	5.95	3.89	2.33	5.80	3.90
	3/400 Average	2.40	4.34	3.25	2.31	5.12	3.54	1.99	4.82	3.32
	USAFA Average	2.53		3.27	1.96		3.50	1.93		3.29

Table 4: Instrument Averages

Division	Course	Outcome #1 Instrument A	Outcome #1 Instrument B	Outcome #1 Instrument C	Outcome #2 Instrument A	Outcome #2 Instrument B	Outcome #2 Instrument C	Outcome #6 Instrument A	Outcome #6 Instrument B	Outcome #6 Instrument C
Engineering	Astro 320*	0.93	0.65	0.52	0.73	0.78	0.60	0.73	0.86	0.65
	Civ Eng 310	0.67	0.50	0.39	0.93	0.75	0.51	0.50	0.55	0.35
	Engr 310	0.93	0.58	0.54	0.33	0.72	0.59	0.47	0.63	0.55
	Astro 410	0.95	0.56	0.55	0.38	0.68	0.60	0.67	0.70	0.62
	Engr 410	0.54	0.44	0.54	1.00	0.75	0.60	0.67	0.54	0.40
Humanities	Engl 311	0.67	0.62	0.55	0.73	0.76	0.60	0.73	0.66	0.48
	Phil 310	0.71	0.66	0.53	0.95	0.86	0.68	0.71	0.75	0.56
Social Sci	Beh Sci 310	0.77	0.79	0.62	0.85	0.95	0.69	0.71	0.93	0.67
	Econ 310		0.73	0.64		0.86	0.68		0.86	0.66
	Law 320	1.00	0.78	0.61	0.89	0.89	0.68	0.67	0.92	0.70
	Law 420	0.83	0.85	0.66	0.89	0.99	0.72	0.78	0.96	0.73
	3/400 Average	0.80	0.67	0.56	0.77	0.82	0.63	0.66	0.76	0.58
	USAFA Average	0.84		0.57	0.65		0.62	0.64		0.57

Table 5: Normalized Instrument Averages

Division	Course	Outcome 1	2	6
Engineering	Astro 320*	0.70	0.70	0.75
	Civ Eng 310	0.52	0.73	0.47
	Engr 310	0.68	0.55	0.55
	Astro 410	0.69	0.55	0.66
	Engr 410	0.51	0.78	0.53
Humanities	Engl 311	0.61	0.70	0.62
	Phil 310	0.64	0.83	0.67
Social Sci	Beh Sci 310	0.73	0.83	0.77
	Econ 310	0.68	0.77	0.76
	Law 320	0.80	0.82	0.76
	Law 420	0.78	0.87	0.82
	3/400 Average	0.68	0.75	0.68
	3/400 Std Dev	0.10	0.11	0.11

Table 6: Composite Averages for Each Course & Each Outcome (three instruments weighted equally)

4.2 Rating Summary

The general philosophy was to rate each course relative to the entire group of 3/400 level courses, not the entire set of USAFA core courses. In addition to presenting some findings about the individual course ratings, Section 4.3 will also address how the 3/400 level compares to the overall USAFA ratings. The final composite ratings (one through seven) were assigned with respect to the course means and standard deviations shown in Table 6 (see Table 7 for the *composite ratings*). The initial ratings were made against the following guidelines (*Note* that for each of the outcomes, there exist different means and standards deviation, as shown in Table 6):

RATING	LOWER BOUND
7	$\mu + 2\sigma$
6	$\mu + \sigma$
5	μ
4	$\mu - \sigma$
3	$\mu - 2\sigma$
2	$\mu - 3\sigma$
1	-

As shown in Tables 4 and 5, Economics 310 was missing data from one of the collection instruments. Thus, after the composite ratings were assigned according to the strict guidelines, subjective adjustment was required to account for the missing data. Economics 310 was missing data from the Course Design Coverage ratings. The 3/400 averages for each outcome were re-

computed to include only the second and third instruments. Econ 310 was then compared to this new average to justify any (or no) changes to the original ratings. Table 7 captures the final ratings, including any subjective adjustments.

Division	Course	Outcome 1	2	6
Engineering	Astro 320*	5	4	5
	Civ Eng 310	3	4	3
	Engr 310	5	3	3
	Astro 410	5	3	4
	Engr 410	3	5	3
Humanities	Engl 311	4	4	4
	Phil 310	4	5	4
Social Sci	Beh Sci 310	5	5	5
	Econ 310	5	5	5
	Law 320	6	5	5
	Law 420	6	6	6

Table 7: Composite Ratings for Each Course and Each Outcomes

These ratings were made relative to the other 3/400 level courses *for each outcome.* A "six" for one educational outcome does not mean that the original rating is the same for a course receiving a "six" for a different educational outcome. Furthermore, these ratings were *not* made relative to the 100- and 200-level courses. In Tables 5 and 6, the 3/400 level averages are presented alongside the USAFA averages to see how the upper level core courses compare to average USAFA core course.

4.3 Findings

This section addresses some key differences and trends found in the Composite Ratings (Table 7), the Composite Averages (Table 6) and the Normalized Instrument Averages (Table 5). First, we will address findings within the 3/400 level. We will also address the 3/400 level relative to all core courses at USAFA.

- Social Science Division upper-level courses scored higher than their counterparts in the 3/400 level.
- Engineering Division upper-level courses scored lower than their counterparts for each of the three outcomes.
- Core substitute courses scored higher than their standard (non-technical) core courses.
- Civil Engineering 310 and English 311 were the only courses to receive below-average ratings with respect to all three outcomes (though English 311 received no 3 ratings).

- The 3/400 normalized ratings for Outcome #2 (Ill-defined Problems) were the highest across all divisions (Tables 5 & 6).
- The design of 3/400 level courses, relative to USAFA, focused more heavily on Outcome #2 (Ill-Defined Problems).
- How each course is *designed* to satisfy each outcome (relative to the USAFA course design averages) is similar to how each course *actually satisfies* the outcomes (relative to the USAFA averages). For example, the 3/400 design average (instrument A) for Outcome 1 is 0.80, while the USAFA average is 0.84. This is similar to what we see for the course "*actually satisfying*" the outcome (instrument C), for which the 3/400 average is 0.56 and the USAFA average is 0.57.
- 3/400 level instructors and course directors tend to overestimate their role in teaching Fundamental Knowledge. This is shown in a simple comparison of the normalized ratings for each outcome, as answered by instructors and course directors in the Course Design Survey (Instrument A) and as answered by the students in the End-of-Course critiques (Instrument B):

	Outcome		
	1	**2**	**6**
Instrument A	0.80	0.77	0.66
Instrument B	0.67	0.82	0.76

The first line gives a normalized profile of the emphasis that the *faculty* feels they are providing each educational outcome. The second line provides the normalized profile of what the *cadets* feel is the contribution of the course to the three outcomes.

5. Summary

This educational assessment was designed to portray the contributions of 300- and 400-level core courses to three of the DF Educational Outcomes. These three outcomes include course contributions to fundamental knowledge, to framing and resolving ill-defined problems, and to developing intellectual curiosity. Data collected during this semester and in previous semesters were consolidated into single course ratings. The purpose of the ratings was to relate each 3/400 level core course to the entire group of 3/400 level core courses. Additional comparisons were made to relate the *group* of 3/400 level courses to the entire set of USAFA core courses. Continued data collection and horizontal analysis of the upper level courses will provide insight into the contribution that these courses make to the cadets' academic and professional development.

300/400-Level Commentary and Critique

1) ***What has this report revealed about the three educational outcomes from the perspective of 300 and 400 level courses?***

As was noted earlier, faculty (even at this level) tend to emphasize integrated knowledge more than either students' ability to frame and resolve ill-defined problems or sustaining their intellectual curiosity. This pattern was first shown in the faculty inputs concerning relative course

emphasis (knowledge was rated 2.40/3.00; framing and resolving received a 2.31/3.00; and intellectual curiosity a 1.99/3.00). This trend was replicated almost exactly in the ratings the team assigned each of the outcomes: knowledge 4.64/7.00; framing and resolving 4.45/7.00 and intellectual curiosity 4.27/7.00. The normalization procedures this team developed and employed made it very clear, however, that students had a somewhat different view. As the comparison of Instruments A & B listed near the end of the report shows, students consider these 3 and 400 level course as contributing relatively more to their ability to deal with ambiguity (framing and resolving) and intellectual curiosity than the courses contribute to students' integrated knowledge.

Using a systematic, relatively objective and somewhat mechanical approach the team was able to identify distinctive differences in course contributions to each of the three outcomes. The relatively broad range of ratings (3.0-6.0) by both this team and the 200-Level team may reflect more about the characteristics of the individual assessors on the team and the assessment process they employed rather than actual increasing diversity in course contributions quality.

2) ***What has the report told us about the contributions of specific 300 and 400 level core courses to the educational outcomes?***

The wide variation in ratings made distinctions between core course contributions very clear. The two Law courses' (320 and 420) contributions to students' development of integrated knowledge were rated 6s significantly above the average assessed contribution assessment of 4.64. In contrast, Civil Engineering 310 and Engineering 410 contributions were rated as 3.0. It should be noted that Engineering 410 is clearly not designed to increase students' knowledge nearly as much as their critical thinking and teamwork (as evidenced by the results of the course design survey--Table 5). Similarly, the relatively low ratings of the two Humanities courses (English 311 and Philosophy 310) may reflect the fact that both these courses are more process than content oriented.

Contributions to the framing and resolving ill-defined problem outcome also ranged from 6.0 to 3.0. Law 420 again led the pack with its rating of 6.0. All of the other Social Science courses as well as Engineering 410 and Philosophy 310 followed close behind with ratings of 5.0. At the other end of the spectrum, Engineering 310 and Astronautical Engineering received ratings of 3.0. An interesting similarity in these courses is that not only do they reside in the Engineering Division, but both are core courses which have technical substitutes (Eng 311 or Astro 320). There is a fairly robust literature that suggests that channeling students into separate skill level tracks contributes little to overall efficiency and runs the risk that teachers' lower expectations of students relegated to the slower tracks will become self-fulfilling prophecies. A "critical thinking" outcome such as the ability to frame and resolve ill-defined problems might be especially susceptible to this phenomenon. These results are not inconsistent with such a hypothesis.

Law 420 again earned top marks for contributions to students' intellectual curiosity. Also similar to the framing and resolving outcome, all other Social Science courses, this time joined by Astronautical Engineering 320, earned positive (5.0) ratings. The Humanities courses and Astro 410 received neutral ratings. The other three Engineering courses Civil Engineering 310, Engineering 310 and Engineering 410) were again rated as being below average (3.0).

Due to a strong positive correlation between the ratings for the three outcomes for each course, the overall results provide an amplified version of the previous discussions. Law 420 is doing a spectacular job of contributing to all three outcomes. All Social Sciences Courses, in fact are doing a very good job as was the core substitute for technical majors (Astro 320). The Humanities courses and Astro 410 received relatively neutral ratings. The other three Engineering courses (Civil Engineering 310 (x=3.3), Engineering 310 (x=3.7) and Engineering 410 (x=3.7)) all received ratings less than neutral. It is again appropriate to consider how well-suited these assessment procedures are to a very non-traditional, process-oriented course such as Engineering 410.

The team's ratings of two other courses were deleted from the original report due to the fact that these courses were not rated by their respective divisional teams (Basic Sciences and Engineering). Both courses were technical substitutes for general core courses: Math 358 and Engineering 311. Both courses were rated relatively highly for their contributions: Math 358 received ratings of "6" for knowledge, "5" for thinking and "5" for curiosity. Engineering 311 received ratings of "5" for each of the three outcomes. It is important to note these ratings here to dispel the impression that only Social Sciences were rated as contributing at a high level to these outcomes at the 3/400 level. It is also useful to review the disciplinary affiliations of the assessors before concluding that this pattern of results reflects common personal or individual biases: 3 of the 4 faculty members on this team, including the team chair, were from the Basic Sciences Division.

3) ***What can be learned from the 300 and 400 Level Horizontal Assessment Team Report concerning the assessment process and practice in general?***

Although there is much less content discussion, the systematic, objective, consensus-building process reflected throughout this report is impressive. The two step process of standardization of scores, comparison and adjustment is particularly commendable. Once again the inclusion of a student on the team seems to have had facilitatory effects. As in the previous (200-level) report, a willingness to assign "below average" ratings makes it much easier to identify opportunities for improvement. Lest the impression be created that this report reflects a "slash and burn" approach, it should be noted that no individual course contributions received less than a 3.0 (slightly below average) rating and only 7 of the 33 ratings (21%) were less than neutral. This overall average rating seems very reasonable.

The fact that this team's assessment accounted for 7 of the 8 significant disagreements between Horizontal and Divisional Assessment Teams is at first a bit disturbing. However, all 7 of these discrepancies occurred with either the Social Sciences or Engineering Divisional teams' ratings. Serious procedural problems (the Social Science team's redefinition (inflation) of the rating scale and the Engineering team's abdication of course ratings to individual department representatives) suggest that these significant differences actually support the validity of the 400-level Assessment Team's ratings.

Note from 3/400 Horizontal Team: The way that we assigned the 1-7 final ratings was done with respect to each outcome. Though this gives us a great comparison of <u>each</u> 3/400 course to the entire set of 3/400 courses for a particular outcome, it does not reflect the relative emphasis

that is placed upon the outcomes by a course. For example, Law 420 received 6, 6, 7 on the final ratings, yet, the normalized composite average were .78, .87, and .82 (Table 6).

There is meaningful information in all the tables that led up to the final 1-7 ratings. I hope the "final reader" of this report recognizes this and doesn't just look at the final ratings. It's clear that Col Porter did not fall into this trap.

Part III
Meta-analysis
Introduction

The previous parts of this report established the institutional context and assessment background, reported this project's charter, and presented and discussed the individual divisional and horizontal reports. Thus far, a number of useful insights have emerged. Meta-analysis is, in some respects, a misnomer. It does not differ qualitatively from the analyses that have been performed throughout this report. Meta-analysis, however, is distinctive in that it occurs at a "higher level." Rather than looking at outcomes, courses and contributions in each separate study, it attempts to take a conceptual step back from the data and consider the larger pattern of results. It attempts to view the reports previously presented as inter-related and interacting components of an overall system: a system of educational outcomes assessment.

Three questions have guided the comments and critique of the seven individual assessment reports: what did each report tell us about the three separate outcomes?; what did each report convey about the courses assessed?; and finally, what did the process tell us about the assessment process and our own ability to assess ourselves and our courses? For the meta-analysis, these same questions will be considered, but in the reverse order. In the last individual report, the 400-level team used a rather mechanical, mathematical formula for assigning initial assessment approximations. The team then allowed subjective evaluations to adjust (or tweak) the initially computed ratings. Similarly, in this report, we have presented raw assessment ratings from each of the separate teams. However, based on the critique of both the assessment ratings and processes, it is not unreasonable to suspect the validity of the ratings varies across the seven teams. If it were apparent that each team's assessment was equivalent and each of the ratings was of equal validity and reliability, no adjustments would be necessary or appropriate. A simple process of averaging the ratings for each course could be used. However, to the extent variation in validity exists, minor adjustments, designed to compensate for identified weaknesses in the original assessment ratings, should provide a more stable data base and thus improve the validity and reliability of the findings from subsequent meta-analysis.

Even minor "tweaking" should not be undertaken lightly, however. The danger of *post hoc* manipulations providing a data base that merely supports preconceived assumptions is significant. For this reason, a systematic and objective review of the products and processes provided by each of the teams is needed before the data base is modified. To the extent that objective weaknesses or biases can be identified, and for which relatively simple mathematical adjustments seem reasonable and appropriate, such adjustments will be made. Where no significant problems are identified or simple mathematical solutions cannot address the weaknesses identified, a simple arithmetic average of the ratings provided by the two teams will be used to arrive at the final assessment rating. Such adjustments should also be made "conservatively" and in such a way as to preserve as much of the original team's intent as possible.

Therefore, the first chapter in this section (Chapter 10) will provide a relatively comprehensive review and evaluation of the process and products of each of the respective

assessment teams. The products (viz., assessment ratings of core course contributions to each of the three outcomes) will be examined first to identify areas of concern for further inquiry. These characteristics include the overall mean of the raw ratings, the range of the ratings, the number of deficient (i.e., below average) ratings, the proportion of "perfect" 7.0 ratings, and the number and proportion of the significant disagreements with other teams' ratings. Additionally, external explanatory information concerning other factors (i.e., hierarchical support or its absence, the team constituency and experience, etc.) will be presented to provide rationale for why particular procedural anomalies or apparent biases may have occurred. Finally, whatever adjustments seem to be appropriate and responsive to the deficiencies identified will be stated clearly and applied systematically to the original raw data base to produce the new data base for subsequent analysis.

The following chapter will then use this data base to examine not only individual course contributions, but also to attempt to identify some of the overarching principles which appear to apply across the Academy's extensive core curriculum. It will also provide the foundation for examining which, if any, curricular characteristics are associated with each of the three individual outcomes as well as an overall general measure of course contributions.

The final chapter in Part III of this report will return to the question of what this project has revealed about the status of the contribution of the Academy's core courses as a whole to the three identified educational outcomes: integrated knowledge, the ability to frame and resolve ill-defined problems, and the development (or retention) of intellectual curiosity. Much of the analysis will simply be a logical extension of the analyses already performed and instances already noted. Some inputs, however, such as the College BASE, and course curricular characteristics which were deemed to be of little value by most of the assessment teams, will be reconsidered with regard to what their results might tell us about the overall core program.

Part IV of this report will then provide a conclusion to this project and brief remarks concerning its implications for the current Unit Self Assessment activities being undertaken by each of the respective academic departments.

Chapter 10

USAFA Phase II Assessment Teams:

Meta-Analytic Evidence of Adequacy?

CHAPTER TEN -

USAFA PHASE II ASSESSMENT TEAMS:

META-ANALYTIC EVIDENCE OF ADEQUACY?

This chapter will address the third question of the interrogative trilogy used to critique the individual assessment team reports. By combining these analyses, patterns of appropriate practice become more salient than in the more narrowly focused preliminary reports. Conceptually, meta-analyses seek to derive general principles by juxtaposing the fruits of separate explorations and establishing a broader comparative framework. It is important to recognize that the adequacy or validity of this report (or any report) is something that must be constructed rather than something that is there waiting to be discovered; assessment validity is contingent upon both inputs and the processes with which it is constructed. Our approach therefore will be incremental: relying on the most objective and observable data to frame the initial problem space, then using more subjective factors to fill in apparent gaps, voids or inconsistencies. The hope is that such an approach will provide a rational basis for combining individual assessment group ratings into a data base worthy of further analysis.

Perhaps the most objective and observable aspects of the assessment teams are the mathematical characteristics of the results (i.e., ratings) they produced. As discussed in Chapter One, teams agreed to use a common 7-point rating scale to quantify the contribution of each of their assigned core courses to each of three educational outcomes: integrated knowledge, the ability to frame and resolve ill-defined problems, and intellectual curiosity. Reducing rich subjective course descriptions and syllabi, faculty opinions and student critiques into a single number obviously involves a loss of information; however, such quantification does allow further relatively objective (and often useful) analyses. In particular, individual ratings can be combined into a single distribution which includes all the ratings of assigned course contributions to each of the educational outcomes. Mathematical characteristics of these distributions such as the mean, the range, and the incidence of certain somewhat unusual ratings (e.g., either perfect or deficient scores) provide an initial indication of the validity of the assessment. Still further information concerning validity can be derived by comparing the ratings provided by a particular team to both the individual contribution ratings given by other teams to the same course contributions as well as the overall distribution characteristics of other teams.

Table 10-1 contains the raw descriptive statistics of the ratings distribution provided by the assessment teams. The number of courses considered by the seven teams varied slightly. Unfortunately, lack of inter-team coordination led to several instances of courses being considered by divisional teams but not by horizontal teams (primarily technical core substitutes) or vice versa (primarily Military Art and Science courses not associated with any of the four academic divisions). Such courses and their uncorroborated ratings were excluded from the final data base (personnel transfers and other priorities made it impossible to reconvene the teams by the time this oversight was discovered) . *Overall average ratings* reflect the sum of all ratings by the team

divided by the total number of ratings given. The overall mean of 5.07 (slightly above average) seems reasonable, but the range of team overall averages of 1.66 (from 6.12 for Social Sciences to 4.46 for the 3/400 HAT) seems large considering the relatively large number of ratings that each team provided (a minimum of 24) and the explicit attempts by most teams to use similar inputs and rating procedures. The next column in the table reflects the overall range of scores provided by each team; the range can be taken as a measure of the team's ability to discriminate contributions across courses and outcomes. The average rating range was 2.16; range variation shows that some teams adopted a relatively conservative approach to assigning ratings (Basic Sciences and the 100 Level HAT) while others adopted a more aggressive approach (200 and 3/400 Level HATs). The *Rating Discrepancy* column shows the average difference in the rating assigned by the team in comparison to the ratings of the same course contributions provided by other teams. Positive scores reflect a team's relative "generosity" to courses in its rating group and negative scores reflect a more critical approach. The overall average discrepancy is the average absolute value of the differences. The next column reflects the number and proportion of significant disagreements with other team's specific ratings; initially this project set aside time for the teams to meet and resolve these discrepancies through discussion, compromise and consensus. However, as the project fell further and further behind schedule; retirements, transfers and competing obligations made these compromise sessions impossible. The last two columns show the number and percentage of ratings in two categories: *perfection* and *deficiency*. As written, the description of a 7.0 rating was intended to limit this rating to truly extraordinary cases -- for the six teams that employed the original rating system, none rated a course contribution as a "7.0." As for "deficient" (i.e., slightly below average) rating, the assumption was that an objective assessment of many courses' contributions to the three different outcomes should yield at least a few deficiencies. Failure to use ratings less than 4.0 (average) might indicate an attempt to provide ratings to impress (or appease) others rather than highlight opportunities for improvement. It might also simply reflect a lack of confidence in either the inputs or the team's own rating process. The way this report is employed is almost certain to influence how many "deficient" ratings are likely to be provided by future assessment teams. Punitive responses to such identified opportunities for improvement would undoubtedly undermine institutional trust and diminish assessment validity in the future. As stated earlier, it is necessary to assume that all those who participated did so with integrity.

The ***Basic Sciences Divisional Team*** provided data which appears very reasonable and consistent with ratings provided by other teams. The overall ratings of 4.78 is just slightly below the overall average of 5.07, but the rating discrepancy score of +.08 is the lowest in absolute terms of any of the teams. The solid consistency of the ratings this team provided is further corroborated by the fact that no significant discrepancies (differences of 2.0 or greater) were noted. The only criticisms that might be applied to this assessment team are the relatively small rating range (1.50) and the team's apparent unwillingness to assign any scores less than average (4.0). The team's report makes it very clear that their concerns about the fuzzy nature of the outcomes and the subjective nature of the cadet critiques led to an intentionally conservative approach to assigning assessment values. Other factors which might be considered as contributing to the success of this assessment team are strong divisional support for assessment (as an application of the scientific method) and strong leadership within the group. Department heads within the division expressed support for educational assessment and identified experienced and capable senior faculty members to represent their respective departments on the divisional

team. The team leader, a very experienced and well-respected senior member of faculty, was a charter member of the Educational Outcomes Assessment Working Group and had been an active participant over the group's two-year development. His selection as a North Central Association accreditation consultant/examiner further reflects the depth and breadth of his knowledge of both education and assessment. Assessment scores from this team will receive unit weighting and be averaged with other ratings assumed to have equal validity.

The ***Humanities Divisional Assessment Team Report*** shows the same pattern of results as well as many of the same strengths and weaknesses as the Basic Sciences Team Report. The Division's overall average (5.08) is nearly identical to the grand average of all ratings (5.07); the team's rating discrepancy score of (+.30) is reasonable; and the single significant difference of 2.00 in rated course contribution suggests general convergence with other teams' ratings. The problems of restricted range and reticence to use deficient scores noted in the Basic Sciences is slightly ameliorated in this report: the team used a full 2.00 range in rating core course contributions to integrated knowledge, but retained a range of 1.5 for the other two outcomes. Once again, however, no course contribution deficiencies were identified. The discussion of ideal versus actual ratings seemed to be a kind of "red herring" which obscured more than it revealed. It also provided a slightly ambiguous framework for subsequent explanatory comments on course ratings. However, the overall pattern of results appears very responsive to the agreed-upon criteria. Once again divisional support for assessment in general and this project in particular was reflected in the strong qualifications and relative seniority of the team members. This team had the advantage of two active and experienced charter members of the Educational Outcomes Assessment Working Group. Recognizing that faculties in the Humanities are often relatively resistant to the necessarily reductionist activities involved in quantitative assessment, the degree of consensus this team achieved is remarkable. Scores from this team will also receive unit weighting and be averaged with other ratings assumed to be of equal validity.

The ***Social Sciences Divisional Assessment Team Report*** reveals several significant deviations from pre-arranged protocols and shared assessment criteria. The relatively large discrepancy average (+.79), the very high proportion of perfect ratings (26%), and the reluctance to identify deficiencies, all suggest caution is appropriate in interpreting these results. The team's initial two-step process of establishing the general advantage of social science core courses over other core courses, then using this evidence to apparently argue that *all* social science core courses *must* be at least above average is of dubious analytic validity. This problem is further exacerbated by the team's decision to abandon the general rating rubric in favor of a completely ill-defined and subjective scale which yielded seven perfect (7.0) contribution scores. It would be nice if the assumption could be made that the team simply had inflated all scores by adding approximately one point to each course's "true" score. However, the report also reveals the team's assumption that more senior courses should receive higher scores by their nature. (The language of the original rubric was specifically designed to eliminate such preconceived biases.) Although evidence might be found to support this hypothesis, asserting it *a priori* and then providing assessment ratings consistent with it (without independent validation or support), seems to be an exercise circularity. It would not be unfair to suggest that senior level interest in this project was somewhat less than in other divisions. The ranking department head announced his retirement during this project and two of the other four department chairs were on sabbatical when the project was initiated. As a result, faculty members on the team had somewhat less

assessment experience and academic seniority than most other divisional teams. Additionally, external circumstances required the team's original leader, and only charter member of the Educational Outcomes Assessment Working Group to withdraw; leaving the team with relatively little assessment experience or expertise. So, should these scores be thrown out altogether or might they be adjusted so they could contribute meaningfully to a common data base? Despite the discrepancies noted, the team did work together to consider the course contribution ratings and the pattern of results seems to be supported by available input data. Due to the team's explicit assumption that more senior level courses deserved higher ratings, most of the inflation seems to have been localized in the higher level social science courses. Translating the scores provided by the team according to a sliding scale would effect the desired statistical changes. Therefore, the following adjustments to raw scores have been made: 7.0 = 6.0; 6.5 = 5.8; 6.0 = 5.5; 5.5 = 5.3 and 5.0 = 5.0. Such a translation reduces the overall divisional average to 5.56 (from 6.12); reduces the range of scores from 2.0 to 1.0; reduces the discrepancy score from .79 to .23, and eliminates all significant disagreements as well as all "perfect" scores. The one problem with this new data set is the severely restricted range. Since variation reflects statistical leverage, the new transformation has reduced the influence of the social science divisional team's ratings; this does not seem inappropriate in light of the process problems noted above.

The ***Engineering Division Assessment Team Report*** also reveals significant problems. Although the raw average overall rating of 5.19 is close to the grand average rating of 5.07, the average discrepancy score of +.91 is even greater than the Social Sciences Team's +.79. This inflationary evidence is further reflected by the fact that 19% (5) of the ratings provided by this team differed significantly from the ones provided by horizontal assessment teams of the same course contributions. To this team's credit no "perfect" ratings were awarded, but once again neither were any deficient scores. One particularly significant process factor is revealed by a closer examination of the team's report. The absence of information about the decisional process coupled with the considerable variety in the depth and rigor of the explanatory comments, suggests that the assignment of ratings was relegated to departmental representatives. These deficiencies do not reflect any lack of effort or commitment by the individuals who rendered the report. They do, however, reflect noteworthy organizational and systemic problems which call into question the validity of some of the ratings. When this project was first considered, the Engineering Team would have had the largest contingency of charter EOAWG members. However, when the assessment project actually began, none of these individuals were available to participate. The alternative leader of the team was a well-respected and very senior member of the faculty; however, he was only able to be present at one of over a dozen faculty assessment coordination meetings and retired just before the first draft of the report was to be submitted. The membership of the other divisional teams contained many relatively senior military and civilian faculty members, the engineering team contained somewhat more junior and less experienced educators. The team's eventual *de facto* leader, a very conscientious and highly motivated individual, was an Assistant Professor who had been at the Academy less than two years.

Unlike the Social Science Team's ratings, the validity of ratings provided by the Engineering Divisional Assessment Team vary by department. This is not inconsistent with departmental representatives being allowed to assess their own courses. General adjustments based on raw scores, as was used for the social science division ratings, thus seem inappropriate.

Another alternative is needed. The relative validity of the ratings provided by the three horizontal assessment teams is supported by their convergence with the ratings provided by the two Basic Sciences and Humanities Division raw ratings and also the "adjusted" Social Sciences scores. If these other scores are used as a benchmark, then the most discrepant divisional assessment scores can be "reeled in." Using an adjustment criterion of 1.0 would necessitate altering only 6 of the 27 divisional scores (Aero 215's score for knowledge, Engineering 310's score for framing and resolving, Engineering 410's score for intellectual curiosity and all three Civil Engineering 310 scores). The resultant rating distribution would have a mean rating of 4.91, a range of 2.00 and the rating discrepancy would be reduced from +.91 to +. 63. The number of significant disagreement would be reduced from 5 to 0 and the distribution would still not contain any deficient or perfect scores. Under the circumstances, adjusting these six ratings seems to be the minimum necessary to protect database integrity and comparability of scores.

The three final reports were provided by the Horizontal Assessment Teams. Although most of the data remains as shown in Table 10-1, the adjustments explained above resulted in changes in some of the distribution characteristics of the Horizontal Teams' ratings. Changes to the Social Sciences and Engineering Divisional raw inputs eliminated all but one of the significant disagreements and also reduced the size of the rating discrepancies. For the 200 Level HAT, the discrepancy was reduced from -.62 to -.48 (i.e., by 23%) and for the 3/400 Level HAT from -1.09 to -.63 (i.e., by 42%). Further discussion of the horizontal team's products and process will use these adjusted values.

The ***100 Level Assessment Report*** is very systematic and conscientious. The overall average is very near the grand average and the small average discrepancy was unaffected by the adjustments to the Social Science and Engineering ratings (only one course from each division was considered by this team). Several possible shortcomings in this assessment are, however, suggested by a review of the characteristics of the rating distribution: no deficient contribution ratings were given and there is once again a relatively small range of ratings. The identification and report of several significant process variables help make this report especially useful as a source of lessons learned for other (future) assessment teams. Although limited in academic rank, team members (including the cadet) had a considerable amount of education and assessment experience. The team was led by a very motivated and capable civilian faculty member who had also been an active participant in educational outcomes assessment from the beginning. Although from a department in the Social Sciences Division, another team member had a doctorate degree in biopsychology, a very technically-oriented degree. The third faculty member had a background in English but was currently working in the office of the registrar with a special focus on assisting "at risk" students. Another significant organizational factor for this team was the availability of a senior, experienced, in-house assessment consultant from the Academy's Center for Educational Excellence. (This individual also provided support and assistance to the other two horizontal teams.) One of the unique developments within this team was the use of ratings varying by as little as .05 rather than .5. A positive effect of this was the ability to make relatively fine distinctions between course contributions (although some might argue the ratings were also even "finer" than the data would support). The capacity to make distinctions at the .05 level may have relieved some of the (appropriate) pressure to use a wider range of ratings and thus necessitate "average" or "slightly below average" ratings. All in all, however, scores from this team are solid and well-supported; they will receive unit weighting and be averaged with other ratings.

The ***200-Level Horizontal Assessment Team Report*** was written by a team of two faculty members and one senior cadet with support provided by the HAT advisor mentioned in the previous paragraph. Its average rating value of 4.70 was slightly lower than the grand average of 5.07. It's wide range (3.0) of ratings, lack of perfect ratings and willingness to assign *slightly below average* ratings of contributions in five instances are all signs of a rigorous and authentic assessment effort. The assessment report itself is very thorough and systematic, providing reasonable support for each of the ratings. Scores from this report will also be applied directly to the data base to be averaged with other validated ratings. This report also provides a valuable exemplar of a systematic approach to educational outcomes assessment.

The ***3/400-Level Horizontal Assessment Team Report*** was also written by a team of relatively junior faculty members drawn from several different academic disciplines and a single cadet interested in and experienced with educational research. Three of the four faculty members were from the Basic Sciences Division and the other was from the Social Sciences Division. The statistical characteristics of the team's distribution of ratings reflect a relatively low average (4.46) and a robust range (3.0) with 7 deficient ratings (21%) being assigned to particular course contributions. Adjustments to Social Sciences and Engineering ratings eliminated all seven of this team's inital significant discrepancies reflected in Table 10-1. Similarly, these adjustments significantly reduced the average rating discrepancy from -1.09 to -.63. The organization and systematicity of the approach taken in this report provides another suitable exemplar for future assessment efforts. The relatively thorough account of the data sources and presentation of the rationale for those used and those excluded is particularly laudatory. The meticulous presentation of all the raw data helps provide strong support for the eventual ratings assigned. Another impressive aspect of this assessment is the "normalization" (actually standardization) and combination of scores from different instruments purporting to measure the same construct. Even better was the team's willingness to reconsider the results individually and then adjust ratings when it was deemed appropriate with respect to logical or subjective criteria. Somewhat like the Social Sciences Divisional Team, this team adopted criteria not included as part of the initial charter rubric. Unlike the divisional team, however, due care was exercised to insure the mathematical criteria developed were consistent with pre-established criteria. This team was also careful to avoid rating inflation and showed no evidence of producing ratings to match preconceived notions about the nature of the courses being rated. Another strong example of a systematic and objective approach to assessment; these ratings will also be taken at face value and combined with others to create the data base for further analysis.

Assessment Team	# of Crs Rated	Overall Avg Rtg.	Ratng Range	Rating Discrp	#/% sigdif	#/% prfct	#/% deficient
Basic Sci	**9**	**4.78**	**1.50**	**+.08**	**0/0%**	**0/0%**	**0/0%**
Humanities	**8**	**5.08**	**2.00**	**+.30**	**1/4%**	**0/0%**	**0/0%**
Soc Sci	**9**	**6.12** *(5.56)*	**2.00** *(1.00)*	**+.79** *(+.56)*	**2/7%** *(0/0%)*	**7/26%** *(0/0%)*	**0/0%**

Engineering	**9**	**5.19** *(4.91)*	**2.00**	**+.91** *(+.63)*	**5/19%** *(0/0%)*	**0/0%**	**0/0%**
100-Level	**12**	**5.19**	**1.60**	**+.12**	**0/0%**	**0/0%**	**0/0%**
200-Level	**12**	**4.70**	**3.0**	**-.62** *(-.48)*	**1/3%**	**0/0%**	**5/14%**
3/400-Level	**11**	**4.46**	**3.0**	**-1.09** *(-.63)*	**7/21%** *(0/0%)*	**0/0%**	**7/21%**
Team averages:	**10**	**5.07** (4.95)	**2.16** (2.01)	**.56** (.40)	**8/8%** (1/1%)	**7/7%** (0/0%)	**12/6%**

Table 10-1

Considering the Big Picture: Evidence for Assessment Adequacy?

This chapter started with a review of descriptive characteristics of the distributions provided by each of the seven assessment teams. These data were then used to re-examine each of the reports and identify facilatory adjustments where faulty assumptions or flawed processes could be identified. It is appropriate to consider the overall extent to which the adjustments improved the initial conditions. Changes to individual reports themselves were discussed within each respective section. Changes to the overall average characteristics are as follows: the overall average team rating was reduced slightly from 5.07 to 4.95; the average rating range was also slightly reduced from 2.16 to 2.01. The average discrepancy between horizontal and divisional assessments, however was reduced by 29% (from .56 to .40) and 7 of the 8 identified significant disagreements were eliminated as were all seven of the perfect, "7.0", ratings. The 12 deficient ratings originally identified remained unchanged. Basically, the adjustments were relatively minor, but to the extent the justifications provided were valid, they should significantly enhance the internal consistency and coherence of the data base and support greater confidence in the results of subsequent analyses. If time and circumstances would have permitted, a much better way to have made these adjustments would have been to have teams consult with one another and reach reasonable compromises. Unfortunately, time and institutional tides precluded this.

Another alternative to the "minimal rational adjustment" approach just completed would have been to simply standardize all the ratings within each team. This approach is purely mechanical and thus, in a sense, impeccably *fair*. However, such an approach rests on the assumption that all teams were rating courses that were making equal contributions to all outcomes with equal amounts of variability. Even a cursory review of the inputs provided to the teams suggests such an assumption is untenable. Distinctive differences between student ratings of academic divisions' core course contributions are readily apparent. Global standardization would have been tantamount to throwing the baby out with the bath water.

In addition to the inspection of overall rating characteristics of the separate assessment teams, it is worthwhile to consider patterns within the overall raw data base. A brief inspection of the overall means and standard deviations for each of the three outcomes for divisional and horizontal teams reveals considerable similarity. Divisional teams' average rated contributions were 5.20 (.62) for knowledge, 5.02 (.66) for thinking, and 5.04 (.56) for curiosity. Horizontal teams' ratings of contributions were 4.84 (.88) for knowledge, 4.81 (.74) for thinking and 4.64 (.82) for curiosity. Horizontal team ratings are slightly (approximately .4) lower and slightly more variable than those of divisional teams. In light of previous discussions these small differences seem to be acceptable.

Inter-rater reliability is the extent to which raters looking at the same phenomenon or characteristic score it the same. High inter-rater reliability does not assure validity (that the rating actually measures what it purports to), however, low inter-rater reliability makes further analyses problematic. In this project, we had seven "observers" (assessment teams) evaluating the contributions of 35 different core courses to each of three educational outcomes (phenomena). Each occurrence was viewed by a divisional team and (simultaneously) by a horizontal team. A comparison of the relationship between these two assigned ratings (as adjusted) yields a measure of internal consistency or coherence of the assessment ratings. Ideally, the best predictor of a horizontal team's rating of Course A's contribution to Outcome X would be the divisional team's rating of that same contribution. A Pearson Correlation Coefficient (r) is the most commonly accepted statistical measure for describing the strength of this relationship. However, the relationship between the strength of covariation between two variables and the value of "r" is not linear; a much better measure of the absolute value of the strength of the relationship is the square of "r", which is the proportion of total variance the two variable have in common. Table 10-2 contains the correlation coefficients, variance and significance of each of the 9 possible comparisons between the ratings provided by two groups (divisional and horizontal) for each of the three outcomes. Statistical significance is the relative probability that the relationship observed might have occurred just by chance. Those relationships indicated with double astrices, had less than a one in one hundred chance of being "random".

Correlations, (Common Variance) and Significance**
of Divisional and Horizontal Teams' Assessments
of Core Course Contributions.

Divisional Team Ratings:	Horizontal Team Ratings		
	Integrated Knwldge	Framing & Resolving	Intellec. Curiosity
Integrated Knowledge	.69 (.48)**	.08 (.01)	.50 (.25)**
Framing & Resolving	.18 (.03)	.67 (.45)**	.44 (.19)**
Intellectual Curiosity	.49 (.24)**	.45 (.20)**	.68 (.46)**

** p<.01

Table 10-2

Table 10-2 reflects a relatively high level of inter-rater reliability. The strongest correlations are found between the two teams' ratings of the same outcomes (.69, .67, and .68

respectively). In absolute terms, the nearly 50 percent common variance is very reasonable, considering the diversity of approaches employed and individual raters involved. Although both teams' ratings of intellectual curiosity were significantly related to other teams' ratings of other outcomes, the strength of these relations was less than half the common variance for ratings of common outcomes (viz., .48 vs .22) (e.g., horizontal and vertical teams' ratings of a courses contribution to intellectual curiosity).

A few final comments are in order concerning what the individual critique of the seven assessment reports tell us about our ability as a faculty to assess our own courses and their respective contributions to three specific outcomes: students' integrated knowledge, their ability to frame and resolve ill-defined problems and their level of intellectual curiosity. Several conclusions seem appropriate. First, teams can produce ratings which approximate common sense notions about the quality and variety of course contributions to all three educational outcomes (i.e., the ratings have relatively high "face validity"). This is not a trivial achievement; several senior faculty members have publicly wondered whether the inclusion of *intellectual curiosity* (as well as other affective outcomes) was relevant and even expressed grave concerns about its measurability. Although relevance is yet to be established, the measurability and apparent variability of core courses' contributions to this outcome seems similar in every way to the teams' ratings of the other outcomes. Intellectual curiosity contributions appear to be no more difficult to assess than contributions to knowledge or critical thinking.

Another conclusion concerns a consideration of the internal and external factors affecting the assessment teams formation, decision processes and subsequently the validity of their ratings. Interdisciplinary faculty teams, without apparent vested interests in course ratings, seemed to produce distributions which had somewhat lower averages and greater variability. This apparent advantage in discriminability might also be attributed to the inclusion of a cadet participant on each of the horizontal teams. Although divisional teams were invited to also include cadet members, none of them did so. Another way of stating this conclusion would be to suggest that highly homogeneous faculty groups, with vested interests in ratings, are likely to have some difficulty in discriminating between courses and their respective contributions to the educational outcomes. In two cases (Humanities and Basic Sciences) these difficulties seem to have been overcome by strong internal leadership and a relatively high level of assessment experience and expertise. Without these assets, such as in the case of the Engineering and Social Science teams, rating problems appear to be inevitable and to some extent insurmountable. It should also be mentioned that support from divisional and departmental heads is necessary to insure the best people are included on assessment teams. These same senior individuals, however, must also be willing to step back once they've selected their best people, and allow the assessment process to run its course free from interference or coercion (real or perceived).

A final comment about leadership will conclude this section. Educational outcomes assessment is relatively new. While it promises to provide valuable information relevant to institutional reform and renewal, its tools and techniques are largely untested and the "best" approach is relatively "ill-defined." An approved solution, guaranteed to yield valid results every time, simply does not exist. Even if it did, personal experience suggests that individuals directed to employ such an approved solution, would be able to find ways to do so that would undermine its validity. Assessment is a fragile process, entirely contingent on trust, good will and the *bona*

fide "best efforts" of everyone involved. The need to sustain institutional trust is even more important than the need to exercise centralized assessment process control. An effort was made to provide adequate resources and support to accomplish the required tasks. Concern for collegiality was given precedence over insistence on process consistency; teams were encouraged to explore alternatives. As demonstrated by the diversity in the team reports themselves, teams felt relatively free to develop their own somewhat unique approaches. To the extent these alternative approaches are "captured" by the explanations within the various reports, *post hoc* analyses and criticisms can be useful in increasing general understanding of educational processes and their assessment. Such an overall *laissez fare* approach is likely to maximize the opportunity for both individual and organizational learning. However, such an approach also involves some inherent weaknesses. One instance where a little more "guidance" would have been very beneficial would have been the early insistence that designated courses be assessed and that the initial rubric be employed. Greater direction and clarity would have insured that this effort yielded even more useful results by including a few more core courses (viz., Military Arts and Sciences courses, technical core alternatives and the new, experimental, and integrative freshman engineering course).

Chapter 11

Core Course Contributions To Three of USAFA's Educational Outcomes: Particulars and Patterns

CHAPTER ELEVEN -

CORE COURSE CONTRIBUTIONS TO

THREE OF USAFA'S EDUCATIONAL OUTCOMES:

PARTICULARS AND PATTERNS.

The table below lists each of the core courses by academic division and its assessed contributions to each of the three educational outcomes. The final, overall column shows the average contribution across all three educational outcomes.

	Basic Sciences					*Humanities*			
Course	**Know**	**F&R**	**IntCur**	**Avg**	**Course**	**Know**	**F&R**	**IntCur**	**Avg**
Bio215	4.25*	4.25	4.75	4.42	**Engl111**	4.95	5.38	5.00	5.11
Chem141	5.35	4.53	4.90	4.93	**Engl211**	3.50*	5.50	5.00	4.67
Chem142	5.33	4.63	4.88	4.94	**Engl311**	4.50	4.75	4.25	4.50
CmSci110	4.70	5.23	4.65	4.86	**FLFr142**	5.00	4.93	5.18	5.03
Math141	5.18	5.00	4.68	4.95	**FLCh132**	5.18	4.95	5.23	5.12
Math142	5.30	4.65	5.00	4.98	**Hist101**	5.70	4.50	5.33	5.18
Math220	3.75*	4.00*	3.75*	3.83*	**Hist202**	5.25	5.25	5.25	5.25
Phys110	5.05	5.40	5.25	5.23	**Phil310**	4.25*	5.25	4.25	4.58
Phys215	5.00	4.25	4.25	4.50					
Avg:	*4.88*	*4.66*	*4.68*	*4.73*	*Avg:*	*4.79*	*5.06*	*4.94*	*4.93*

	Engineering					*Social Sciences*			
Course	**Know**	**F&R**	**IntCur**	**Avg**	**Course**	**Know**	**F&R**	**IntCur**	**Avg**
Aero215	5.00	4.25	4.75	4.67	**BSci110**	5.25	5.85+	5.75+	5.62+
Astro320	5.50	4.00*	5.50+	5.00	**BSci310**	5.25	5.50	5.25	5.33
Astro410	5.50	3.50*	4.50	4.50	**Econ221**	4.25	5.50	4.75	4.83
Engr310	5.00	3.50*	3.50*	4.00*	**Econ310**	5.50	5.25	5.25	5.33
Engr410	3.50*	4.50	3.50*	3.83*	**Law320**	5.75+	5.40	5.25	5.47+
ElEng215	5.50	4.75	4.50	4.92	**Law420**	6.00+	6.00+	6.00+	6.00+
ElEng231	6.00+	5.75+	4.75	5.50+	**Mgt210**	5.00	5.25	5.00	5.08
EMech120	5.73+	4.88	4.70	5.10	**PSci211**	5.75+	5.75+	6.00+	5.83+
CEng310	3.50*	4.50	3.50*	3.83*	**PSci212**	5.50	5.50	5.50+	5.50+
Avg:	*5.03*	*4.40*	*4.36*	*4.60*	*Avg:*	*5.36*	*5.56*	*5.42*	*5.45*

* more than one standard deviation below grand mean
\+ more than one standard deviation above grand mean

Considering the extensive discussion of the assessment ratings and processes contained in previous chapters, only a brief synopsis of these results will be included here.

Contributions to the educational outcomes from courses in the ***Basic Sciences*** were generally very consistent; a slight emphasis on developing students' knowledge over either their critical thinking or intellectual curiosity was apparent in most courses in this division. Computer Science 110 and Physics 110 were, however, notable exceptions to this tendency. Physics 110 appears to be the strongest overall contributor to the three educational outcomes assessed in this project. Biology 215's contribution to integrated knowledge was just below one standard deviation below the grand mean for all core courses. This is somewhat disappointing in light of its placement in the Basic Sciences core curriculum after math, computer science, chemistry and physics. Biology has not traditionally been presented as a platform for integrating and applying principles and concepts from other basic sciences, but perhaps explicit consideration of such connections might be worth exploring. One course that was given relatively low scores across the board was Mathematics 220, Statistics. Designed primarily for non-technical majors, Math 220 may be an example of the relative disadvantages of "tracking" students (i.e., assigning students to sections based on their expected level of performance or ability). The major disadvantage of this approach is that teachers' lowered expectations about the potential of students in the "slow" sections inevitably become self fulfilling prophecies. Although not assessed by the Basic Sciences Divisional Team, Math 357 and 358 (the technical substitute for Math 220) regularly receive much more positive ratings from students than Math 220.

Unlike the Basic Sciences, most ***Humanities*** core courses placed greater emphasis on critical thinking ability and intellectual curiosity than on the development of integrated knowledge. English 211, Masterpieces of Literature, and Philosophy 310, Ethics, in fact, appear to invest so much emphasis in teaching students the thinking process, that the contributions of these courses to cadets' integrated knowledge was assessed to be more than one standard deviation below the grand mean. History 101, Modern World History, is the most notable exception to this tendency. It should be pointed out that the two foreign language courses were selected by the Divisional Team (and also assessed by the 100-Level Horizontal Assessment Team) as being generally representative of the wide array of language courses from which cadets may choose. History 202, Introduction to Military History, is noteworthy for its balanced emphasis and demonstrated success in contributing to all three of the educational outcomes. Although, none of the three Military Arts and Sciences Courses were assessed, a brief review of cadet critique data from these courses suggests that military subject matter is seen by cadets as being understandably relevant and applicable.

Social Sciences core courses made strong and balanced contributions across the board, with just slightly more emphasis on the framing and resolving outcome than on the other two outcomes. Several courses were distinguished by ratings more than one standard deviation above the grand mean for all core course contributions. Law 420, Law for Commanders, earned top marks in every category. Political Science 211, Politics and American Government, and Behavioral Sciences 110, General Psychology, were also rated as being overall exceptional; both courses placed slightly more emphasis on intellectual curiosity and students' ability to frame and resolve ill-defined problems. Law 320, The American Legal System, was rated as emphasizing integrated knowledge, while Political Science 212, International Politics and U.S. National

Security, placed equal emphasis on all three outcomes; both were assessed as also being overall top contributors. In a field of very strong contributors, Economics 221's, "slightly above average" contribution scores appear to be deficient; yet compared to overall core course means they are near average. Three courses within the Social Sciences Division deserve particular attention. Econ 310, Behavioral Sciences 310, and Law 420 were collectively the Social Science Option. Students majoring in either Basic Science or Engineering Programs could actually choose of these three courses; students in Humanities or Social Sciences programs were required to take all three. The average contributions of these courses were well above the grand mean (knowledge 5.58 (+.56), ill-defined problems 5.70 (+.78), and curiosity 5.67 (+.83)). Although not quite as great, these courses also showed an average contribution higher than the overall social sciences core course contributions (knowledge +.22, ill-defined problems +.14 and curiosity +.25). In this instance curricular choice appears to have not only been good for students, but for these three courses as well.

Assessed contributions within the ***Engineering*** Division appear to be the most variable. This might have been due to the relative inexperience of raters, but the worst of these inconsistencies were adjusted as described in the previous chapter. The remaining differences might be considered as reflecting something substantive about the courses themselves. On the positive side Electrical Engineering 231 (the "technical" alternative to the core requirement) received distinctive ratings for its contributions to integrated knowledge, students' ability to frame and resolve ill-defined problems and overall contribution. Astro 320 (the "technical" alternative to the Astro 410 core requirement) received a rating for contributions to intellectual curiosity more than one standard deviation above the grand mean. Although not assessed by the Horizontal Assessment Team, raw data pertaining to the new, integrative, problem-based Engineering 110Z course suggest that it probably would have also received very high marks. On the other hand, three courses received scores relating to two or more outcomes that were more than a full standard deviation below the respective grand means. Civil Engineering 310, Air Base Design and Performance, and Engineering 410, Engineering Systems Design, received the lowest overall contribution scores. A review of the research in higher education reveals that student attitudes toward courses often intensify or fade over time but do not change polarity (i.e., shift from negative to positive or vice versa). However, a host of anecdotal evidence suggests Engineering 410 may be an exception to this general rule. There seems to be little similar belated adulation for the Civil Engineering course, however. Engineering 310, Energy Systems (another technical core course for which there is an alternative "technical" substitute, Eng 311) also received very low scores for both framing and resolving ill-defined problems and intellectual curiosity. Both Astronautics courses (the high tech and the low tech versions) received scores more than one standard deviation below the grand mean for contributions to cadets' ability to frame and resolve ill-defined problems, although Astro 320 (the "technical" alternative) received distinctively high scores for its contributions to cadets' intellectual curiosity. This overall pattern of results is consistent with the predicted negative influence of differential faculty expectations and self-fulfilling prophecy aspects of skill "tracking" discussed in relation to the Math 220 course. The mixed message sent to cadets by requiring them to take it, but only offer it as a "watered down" non-technical version, seems to set the stage for classroom conflict and frustration for students and faculty alike.

While far from perfect, contribution scores do appear to tell a coherent and comprehensible story. Core courses at the Air Force Academy contribute to cadets' acquisition of a body of integrated knowledge, development of the ability to frame and resolve ill-defined problems and the sustenance of their intellectual curiosity. There is variation from course to course in the relative emphasis placed on these three outcomes as well as the assessed quality of those contributions. These ratings distinguished several core courses for their overall contributions to these three outcomes: Law 420, Political Science 211, Behavioral Sciences 110, Electrical Engineering 231, Political Science 212 and Law 320. In contrast, assessment ratings received by several courses give cause for concern and may warrant further inquiry. Mathematics 220, Civil Engineering 310, Engineering 310, and Engineering 410 all received overall contribution scores more than one standard deviation below the grand mean.

Chapter 12

What's Been Learned About The Three Educational Outcomes

CHAPTER TWELVE -

WHAT'S BEEN LEARNED ABOUT THE

THREE EDUCATIONAL OUTCOMES

In this chapter each of the three educational outcomes assessed as part of this project will be reconsidered in light of the information garnered from this assessment project. It's important to recognize from the onset that the three outcomes which were the focus of this project are only part of a larger set of seven outcomes developed through an iterative, consensus-building process by senior faculty members over the past several years. One of the outcomes, integrated knowledge, served as a divisional focus for a full day Permanent Professor Off-site Workshop in 1993. The other two outcomes were examined as part of Phase One of the Educational Outcomes Assessment Working Group Charter. This involved literature reviews extensive discussion, development of descriptive criteria at three levels of performance (viz., *excellent, satisfactory* and *deficient*), as well as various ideas concerning tools and techniques for course and program assessment. In addition to the value of these products, the value of developing in-house faculty assessment expertise should also be noted; teams led by individuals who had worked through the above assessment activities from the beginning, produced superior assessment reports. Before considering implications of the Phase II project, it's appropriate to provide a brief review of the work that had already been accomplished concerning each of these three outcomes. Each outcome will be presented verbatim with the explanatory comments extant at the time of the charter for this project. Results from the College BASE will be discussed for the integrated knowledge and ill-defined problems outcomes. EOAWG assessment comments from their Phase I: Initial Report as well as descriptions of the *excellent* level of performance for the framing and resolving and intellectual curiosity outcomes will also be presented.

OUTCOME 1: Officers who possess a breadth of integrated, fundamental knowledge in the basic sciences, engineering, the humanities, and social sciences, and a depth of knowledge in an area of concentration of their choice.

Breadth of fundamental knowledge *in these four domains is the essential foundation of intellectual competence and adaptiveness in a complex and changing world. More than knowing mere facts, integrated, fundamental knowledge refers to competence in solving basic problems characteristic of different disciplines and in discerning key relationships among disciplines. This knowledge base must also provide graduates with an awareness of the technological, social, political and economic complexities of international as well as domestic issues.* ***Depth of knowledge*** *in an optional area helps develop that awareness and the abilities described in the remaining outcomes.* (It should be noted here that although the Phase II charter included this outcome, the actual assessment dealt primarily with breadth, the core's contribution, rather than the depth of knowledge, assumed to be supported more directly by the academic majors programs.)

As mentioned previously, this outcome was one of the first to be dealt with by senior faculty members. Divided by academic divisions, permanent professors (department heads) worked together to provide one-page elaborations of how fundamental knowledge might best be described within their respective academic divisions. This activity generated many useful and enlightening conversations and insights; all four divisions reached consensus on relatively concise and distinctive statements of what constituted *fundamental knowledge* within their respective academic divisions. These statements were provided to the respective assessment teams as part of their original inputs.

Unlike the two other educational outcomes, *knowledge* has long been recognized and understood as an appropriate measure of academic performance. Previous studies, such as the College BASE examinations discussed in Chapter 2, provide information relevant to the overall level of cadets' knowledge in many domains covered by the core curriculum. Because this instrument was originally used in such a broad manner (viz., given only to representative samples of freshmen at the start of their first semester and seniors at the start of their senior year in a cross-sectional design), its results were not seen as being particularly relevant by any of the individual assessment teams. However, at a very general level, cadets' accumulation of knowledge over their four years at the Academy is shown by significant increases in the proportion of students scoring in the highest category on the various disciplinary subscales of the College BASE examination. For example, the College BASE includes six categories under English: reading critically, reading analytically, understanding literature, writing as a process, conventions of written English, and a writing exercise. On average, 22% of Academy freshmen scored in the "High" classification category; in contrast 34% (an increase of 12%) of the seniors were scored in this highest category on the same tests. Another Humanities core course sequence is History. Although an average of 24% of freshmen scored in the "High" category for understanding the significance of U.S. and world events, 46% (+22%) of senior cadets scored in the "High" category. This improvement is also likely to include contributions from several Social Science courses such as Political Science and Economics.

Several of the College BASE scales also relate directly to the Basic Sciences. On average, freshmen cadets' knowledge of mathematics is much stronger (than their knowledge of English and the Humanities); on average, 49% of freshmen scored in the highest category on the 7 subscales in Mathematics on the College BASE; on average 58% of the seniors scored in the "High" category, a net gain of 9%. The College BASE has three subscales relating to laboratory and field work: observation/experimental design; laboratory/field techniques and interpreting results. On average, 46% of freshmen and 53% of seniors earn the "High" performance category. However this small (7%) improvement conceals a somewhat disturbing interaction: design scores actually decrease by 20% (from 47 to 27% in the "High" category) while technique scores increase by 27% (from 53 to 80%). These results might be taken as evidence that cadets actually do develop a "plug-and-chug" mentality that emphasizes mechanics and precision at the expense of the broader theoretical perspectives required for experimental design. The College BASE also includes a subscale concerning the Physical Sciences (primarily Chemistry and Physics); 53% of the freshmen and 65% of the seniors score in the "High" category, an increase of 12%.

The College BASE subscale for Life Sciences includes both Biology (a Basic Science at USAFA) and Psychology (a Social Science at USAFA). Only 18% of freshmen score in the

highest category but 63% of the seniors do, an increase of 45%. Other Social Sciences are also covered by College BASE subscales. The Political/Economic Structures subscale also shows a substantial, 28%, increase from 42% of freshmen to 70% of seniors scoring in the top category. Social Science Procedures increase from 34% freshmen "Highs" to 45% senior "Highs" (+11%). Students scoring in the "High" category in Geography, which is not currently taught as part of the core, increases slightly from 66% for freshmen to 70% for seniors.

Few undergraduate colleges and universities have undergraduate engineering core course requirements, so the College BASE has few subscales that relate directly to fundamental engineering knowledge. However, most of the subscales within mathematics seem to be somewhat relevant. Subscales, followed by the average proportion of freshmen scoring in the "High" category, proportion of seniors in the "High" category and net change are as follows: Practical Applications (42%; 77%; +35%); Properties and Notations (61%; 32%; -29%); Evaluating Expressions (34%; 60%; +26%); Equations and Inequalities (61%; 60%; -1%); 2 & 3 Dimensional Figures (45%; 47%; +2%) and Geometrical Calculations (58%; 67%; +9%). Similarly, it seems possible the improvement in Laboratory/Field Technique (+27%) as well as the observed decrease in Observation/Experimental Design performance (-20%) may be related to content covered by Engineering core courses as well as the way this content is conveyed. However, these particular results reflect so much variability, it's hard to draw even tentative conclusions.

To summarize these results: freshmen arrive with remarkable math, science and engineering knowledge. Seniors show moderate increases across the board; about 10% more score in the "High" category then do freshmen. Life Sciences appear to be the area of greatest increase in knowledge. Freshmen appear to be somewhat weaker in English and Humanities knowledge than Math and Science knowledge. Although English and Humanities knowledge shows more progress, they remain relative soft spots in senior cadets' epistemological framework. The largest gains in knowledge (as reflected by increases in the proportion of students scoring in the "High" category) appear to occur in the social sciences where many increases of 20% or more can be found with no decreases. Over 50% of seniors score in the "High" category in most areas in Math, Science, Social Sciences, and Engineering. Approximately 35% score in this category in English and more achieved top Marks on the actual essay portion of the College BASE.

OUTCOME 2: Officers who can frame and resolve ill-defined problems.

***Ill-defined** problems are ambiguous, interactive and ever-changing. **Framing** means constructing a working model, and revising it based on feedback. **Resolving** means that an ill-defined problem is never solved for good; rather it is solved again and again (re-solved) as the problem is framed again and again; and each successive solution is more refined (resolution).*

This Educational Outcome and its explanatory paragraph are listed above. The assessment commentary and rubric defining *excellent* performance as developed by the original Educational Outcomes Assessment Working Group are shown below.

"In assessing student skills in this area, it is important to recognize that the problem should be viewed from the *student's perspective*, not necessarily from the perspective of experts in the

field or even the faculty members evaluating performance. This suggests that different types of problems will be appropriate for assessing students' general ability and their abilities within their chosen academic specialty. It is important to point out that it is the solution process that must be assessed, not just the solution. In fact, a student who had already learned 'the approved solution' from independent reading, might be less likely to demonstrate a high level of framing and resolving skill. The purpose of constructing ill-defined assessment tasks for students is to assess our graduates' ability to recognize and contribute to the resolution of real-world dilemmas they are likely to face in their future as Air Force officers.

In addition to the definitions contained in the outcome itself, it is important to point out that ill-defined problems have no single, absolute solutions. However, solutions to these problems are more than a matter of opinion or mere preference; viable criteria exist for evaluating solution quality. Ill-defined problems frequently contain extraneous information and often lack necessary data. To provide meaningful assessment, tasks to evaluate students' skills must be carefully tailored to challenge students but not overwhelm them. Assessment of both individual and group ability to frame and resolve ill-defined problems should be undertaken across all four academic years." (EOAWG Phase I: Initial Report, pg. 14)

The EOAWG Phase I: Initial Report also contains a set of descriptors relating to three levels of performance relevant to students' ability to frame and resolve ill-defined problems. Only the elaboration of the *excellent* category is listed here; elaborations of performance at the *satisfactory* and *deficient* levels used a consistent process framework and can be found in the original report. Characteristics of the performance of students with *excellent* ability to frame and resolve ill-defined problems were:

"- identifies the most important ill-defined aspects of problems as well as the general 'ill-defined' problem nature
- keenly aware of personal perspective and biases and compensates effectively
- also aware of relationship between present problem and context in which it is situated
- uses goal, mission and other 'ultimates' to structure problem space effectively
- systematically works through problem, often makes multiple passes through the problem space as conditions change in order to assess consequences of changes or alternatives
- unsuccessful attempts regularly used to better understand problem and solution process
- generates rich variety of alternatives; tests them objectively and selects rationally
- uses general principles and fundamental concepts to frame problem space and as solution tools
- provides reasonable and substantive justification for assumptions and choices
- manifests an appropriate level of confidence and commitment to eventual solution."
(EOAWG Phase I: Initial Report, pg. 15)

The EOAWG Phase I: Final Report, Chapter 1 reports on the development of a particular instrument designed to assess students' ability to frame and resolve ill-defined problems. Half the seniors in the class of 1995 completed this instrument; results showed that students did vary considerably in their ability to frame and resolve ill-defined problems and that raters could discriminate between five categories of performance very reliably. Slightly over 60 percent of the cadets scored in the satisfactory category or higher, 25 percent in the category between satisfactory and deficient; and about 15 percent in the deficient category. However, the results of

this study showed only very small correlations ($r = .20$; $r^2 = .04$) with GPA and chosen academic major (although students majoring in the Humanities tended to score slightly higher than students majoring in programs in other academic divisions). Similarities with the Reflective Judgment Model and its measurement of distinctive levels, as well as encouraging positive correlations with the GPAs of particular academic majors, supported the validity of the instrument. Although considerably more validation and calibration work needs to be done with this particular instrument, these preliminary results appear sufficient to warrant concern and further investigation of the Academy's effect on students' ability to frame and resolve ill-defined problems. College BASE results suggesting cadets' improvements in the ability to apply techniques are offset by an apparent decrease in their ability to "design" experiments provide additional cause for concern.

The College BASE discussed in Chapter 2 and again revisited in the discussion of cadets' level of integrated knowledge, also contains a section directly relating to students' thinking competencies. Three discrete levels of competencies are assumed to tap development at different cognitive levels. *Interpretive Reasoning* is the cognitive process students use to understand information in their own terms. *Strategic Reasoning*, the next level of cognitive development, involves processes such as defining, comparing, classifying, analyzing and inferring or deducing. The College BASE labels the highest level of reasoning as *Adaptive*. This type of reasoning deals with understanding information in context and being able to create and employ conceptual models of processes to make viable predictions. *Adaptive Reasoning* seems similar to the USAFA's rubric developed to describe an *excellent* ability to frame and resolve ill-defined problems. *Strategic Reasoning* contains many of the same components as our elaboration of satisfactory framing and resolving skills and *Interpretive Reasoning* seems to define the borderline between the satisfactory and deficient categories. On average 97% of freshmen scored in the "High" category for *Interpretive Reasoning*, 66% for *Strategic Reasoning* and only 11% for *Adaptive Reasoning*. As for seniors, on average, 97% scored in the "High" category for *Interpretive Reasoning*, 97% in the "High" category for *Strategic Reasoning*; and 22% for *Adaptive Reasoning*. While not overwhelming, these data suggest that nearly all of the 34% of cadets not scoring in the "High" category for *Strategic Reasoning* (or Satisfactory on the framing and resolving performance criteria) develop the skills needed to score in the top category prior to graduation. Approximately 1 in 9 freshmen score in the top tier on the most complex and demanding thinking tasks; by the time they become seniors, slightly more than 1 in 5 can achieve the "High" standard in this category.

OUTCOME 6: Officers who are intellectually curious.

Beyond possessing knowledge and having abilities to put that knowledge to active use, graduates of the Academy must be inclined to do so. We want to develop an ***attitude of intellectual curiosity*** *in our graduates that predisposes them to lifelong learning.*

The EOAWG also added several paragraphs of discussion to this definition and a brief explanation: "Intellectual curiosity is an affective orientation; it reflects how an individual feels about the processes of thinking and learning. Those with a high level of intellectual curiosity are eager to understand; they are inclined to solve problems for the sake of finding an answer rather than receiving an extrinsic reward. Those with true intellectual curiosity are motivated by intrinsic factors to accomplish academic tasks. Rarely do individuals display the same level of intellectual

curiosity toward all tasks in all domains; some things are likely to interest each of us, but few are interested in everything. Typically, the degree of intellectual curiosity varies from one discipline to the next. Work within the academic discipline in which a student has chosen to major is the most likely place to find the highest level of intellectual curiosity. As necessary as such specific curiosity is, it is not sufficient. It is also important for our graduates to develop a more general curiosity which can provide the framework for lifelong learning and the integration of their knowledge and experience across many intellectual domains.

Because intellectual curiosity is a motivational characteristic rather than a specific skill or mastery of a particular body of knowledge, special emphasis is needed to understand its assessment. In proposing *excellent, satisfactory* and *deficient* levels of performance, it is important to note that we are describing characteristics or attributes of our graduates not the graduates themselves. Our graduates' level of curiosity in particular academic domains may provide a great deal of useful information about our curriculum and teaching techniques; it is unlikely to tell us much about the individual interests of students we brought to the Academy. Blaming graduates for a lack of interest in a particular academic area would be to turn the entire assessment effort upside down. To a great extent, the level of intellectual curiosity our graduates display is a reflection of the educational experiences they've encountered as students and not arbitrary choices on their part. To interpret deficiencies in this area as the fault of cadets would be to invite the imposition of intellectual "political correctness" which would only serve to alienate both students and faculty. (EOAWG Phase I: Initial Report, pg. 16)

The working group described *excellence* on this outcome as becoming an independent/autonomous learner, and went on to elaborate what they considered to be key characteristics of such excellence:

"- Demonstrates interest in material and study beyond requirements; raises questions that take learning beyond immediate, external goals; consistently exceeds minima
- Reflective and challenged by external world; eagerly seeks out inconsistencies which may reveal opportunities for learning and systemic improvement; excited by opportunities to explore diverse domains or perspectives
-Recognizes and appreciates both the intrinsic and extrinsic value of knowledge
- Fascinated by intellectual challenges across academic domains; thrives on ill-defined problems
- Independently synthesizes information from previous experiences; continually works to discover connections between academic disciplines and apply academic concepts to real life situations." (EOAWG Phase I: Initial Report, pg. 17)

Several studies in the EOAWG Phase I: Final Report addressed factors influencing the intellectual curiosity outcome. Chapter 2 of the EOAWG Phase I: Final Report contains a brief account of a student research project involving both a student survey and several subsequent focus groups. Several distinctive themes emerged from the focus groups: "1. Providing students with real choices is essential... 2. Faculty need to be better trained to understand education... 3. There is a clear difference in the kind of activities that contribute to student learning and those that evaluate student knowledge... 4. Attitudes are important..." (EOAWG Phase I: Final Report, pg. 14). Two other studies in the report used student critiques to assess contributions to the intellectual curiosity outcome: departments whose faculty members claimed to encourage

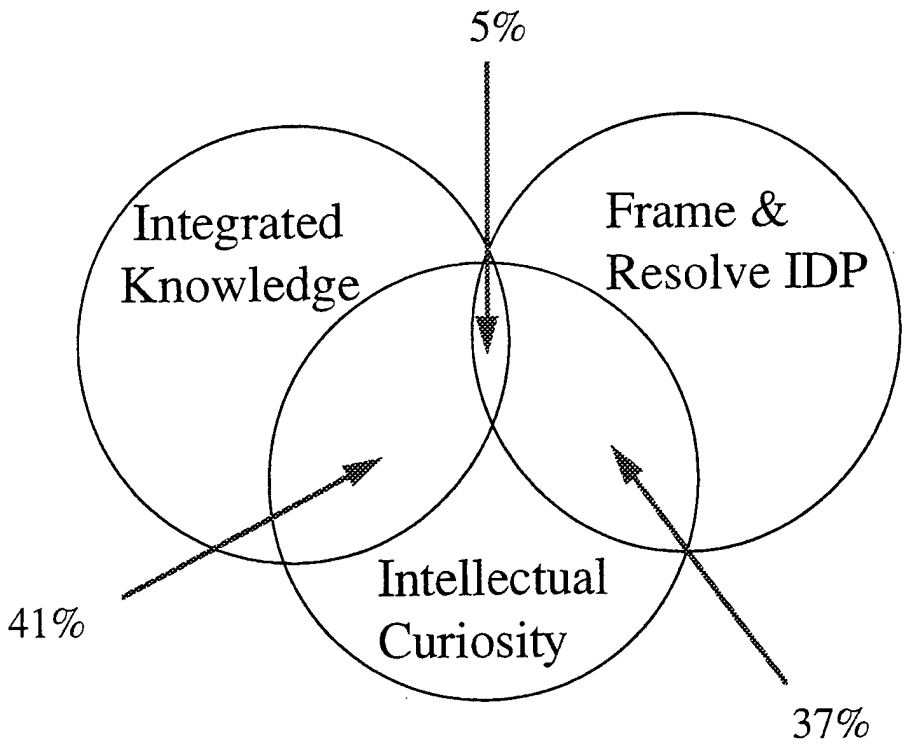
students to challenge ideas in the classroom, made greater use of small groups and also provided more lectures were rated by students as contributing the most to their intellectual curiosity (p. 41). In a study of over 110 individual faculty members, those who placed relatively equal emphasis on students' attitudes and knowledge were found to receive much higher ratings than did faculty members who placed greater emphasis on knowledge than on attitudes (p. 55).

In addition to these EOAWG activities, the whole faculty has been involved in colloquia concerning the intellectual curiosity educational outcome and its assessment. In particular, questions of definition, measurability, influences, related classroom techniques, and consequences were all discussed within departments, interdisciplinary faculty groups at various levels, and finally in a forum open to all faculty members. Issues, insights and common ground discovered through this process will be published as a separate report. However, it is important to mention this activity here because it shows that the institutional climate is becoming more conducive to open dialog concerning the appropriate role for this important, although still somewhat controversial, educational outcome.

Phase II Data

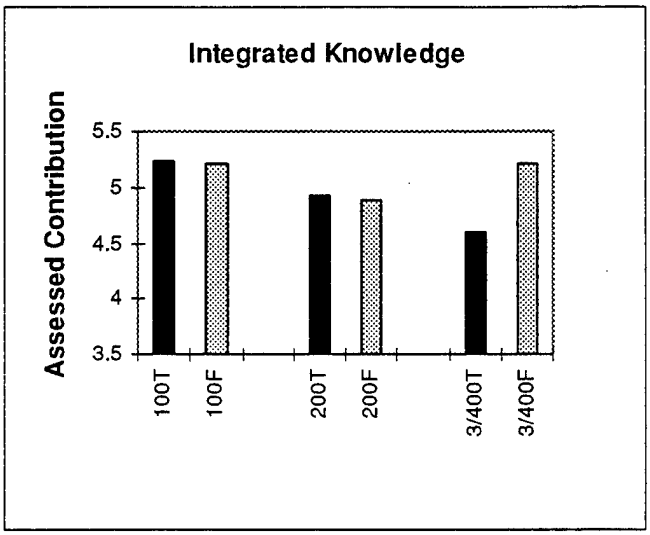
Overall average core course contributions to all three outcomes were rated as being significantly above neutral. The overall mean scores, standard deviations, and corresponding t value comparisons to neutral were as follows: *integrated knowledge* (5.02, sd = .69, t = 8.62); *framing and resolving* (4.92, sd = .64, t = 8.38); and *intellectual curiosity* (4.84, sd = .64, t = 7.65). Although there is still considerable room for improvement, these results show that as a whole, core courses provide a relatively balanced emphasis and as a whole make moderate contributions to each of these three educational outcomes.

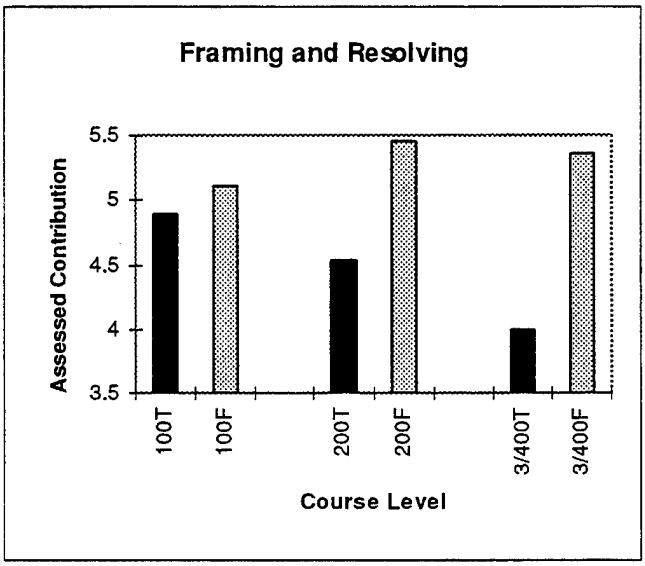
Ratings contained in this report also allowed an examination of the relationship between the three outcomes. Previous factor analyses of course critique data suggested that students saw all educational outcomes being more closely related to one another than they were to a range of other items relating to either course or instructor characteristics (viz., outcome contribution ratings clustered together to form a single *factor*). However, in this study, student critiques and surveys were only a part of the information available and the assessment teams were primarily comprised of experienced faculty members. Their pattern of assessment ratings across courses thus provides "a fresh look" at the nature of these three educational outcomes as they are combined within the 35 core courses. Once again Pearson Correlation Coefficients were calculated to identify the relationships and proportions of common variance between assessment ratings of each educational outcome. The rating a course received for Intellectual Curiosity was closely and positively related to the rating it received for contributing to both *integrated knowledge* ($r = .64$, $p < .01$) and *framing and resolving* ability ($r = .61$, $p < .01$). However, the relationship between course contributions to integrated knowledge and framing and resolving ability was not significant ($r = .22$, $p > .05$). To make this point even more clearly Figure 12 - 1 contains a Venn diagram with the approximate areas of common variance between the three outcomes:

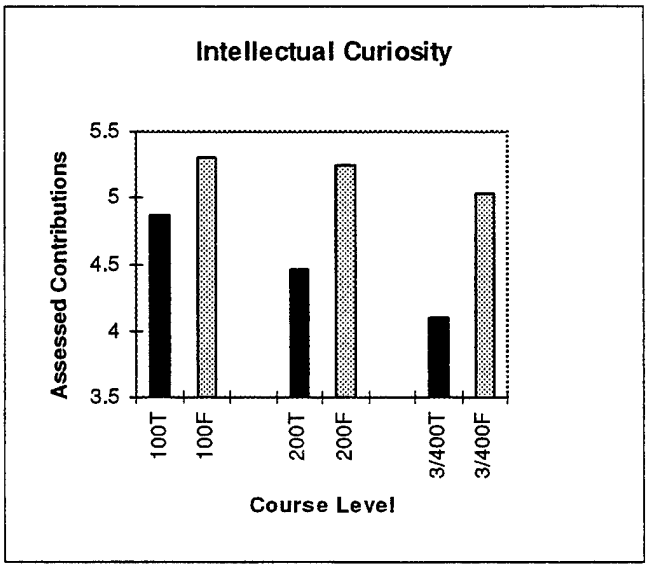


Although the relationship between framing and resolving ill-defined problems and developing integrated knowledge is not significant it is positive. One interpretation of these results, which is consistent with other studies and findings discussed in this chapter, is that contributions to intellectual curiosity are like putting money in a classroom "bank". The more that is put in initially, the more intellectual interest will accrue so that more work can be done throughout the semester. Once curious, cadets can then be enticed to satisfy this curiosity by learning things (increasing knowledge) and by developing skills (framing and resolving ability). The fact that the correlation between these two latter outcomes is positive (but admittedly small) suggests that content does not necessarily have to be reduced in order for students to acquire critical thinking skills. In fact, it appears early investment in students' intellectual curiosity (i.e., showing how the material is relevant, interesting or even exciting) can support high levels of contribution to both the knowledge and thinking outcomes.

A vernacular method of dichotomizing the core curriculum involves the distinction between Basic Science and Engineering courses ("Techys") and Humanities and Social Science courses ("Fuzzys"). The following graphs were derived by collapsing data for "techy" and "fuzzy" courses at three levels (100, 200 and 3/400). The goal was to conduct a visual examination of the data to ascertain patterns in the contributions to outcomes by these two broad categories of courses across cadets' four years at the Academy.


Integrated Knowledge
Assessed Contribution
5.5
5
4.5
4
3.5
100T
100F
200T
200F
3/400T
3/400F


Framing and Resolving
Assessed Contribution
5.5
5
4.5
4
3.5
100T
100F
200T
200F
3/400T
3/400F
Course Level


Intellectual Curiosity
Assessed Contributions
5.5
5
4.5
4
3.5
100T
100F
200T
200F
3/400T
3/400F
Course Level

The graphs above suggest that "fuzzy studies" were seen by the assessment teams as contributing significantly to each of the three outcomes across the three levels at a consistent and relatively high level. At the 100 level, "techy" courses appear to offer roughly commensurate contributions to integrated knowledge and framing and resolving ability but contribute slightly less than "fuzzies" do to intellectual curiosity. At the 200-level, "techy" courses were rated as continuing to contribute to integrated knowledge, but contributions to cadets ability to frame and resolve ill-defined problems are nearly a full point below the contribution of "fuzzy" courses. The greatest relative deficits, however, appear to occur for the technical courses taught at the 3 and 400 level. The fact that students have already selected an academic major at this point may predispose students to resist course demands in disciplines different from their chosen field. As discussed under the course contribution section, the creation of "fuzzy versions" of technical core courses may serve to exacerbate this problem. When students' initially resistant attitude is combined with the lowered expectations of technical faculty tasked with teaching non-technical versions of their favorite subjects, the potential of adversarial relations between students and teachers would seem to be relatively high.

Further evidence of the differential contributions of technical and non-technical versions of core courses can be found by examining expected and actual grades given in the courses. The table below lists the expected and actual grades given at the end of the Fall 1996 semester for four pairs of courses.

Non-Technical Version				Technical Version			
Course	Exp.GPA	Act.GPA	Diff.	Course	Exp.GPA	Act.GPA	Diff.
Mth220	2.84	2.75	-.09	Mth357	2.93	2.98	+.05
EE215	2.82	2.72	-.10	EE231	3.04	2.94	-.10
Eng310	2.86	2.78	-.08	Eng311	2.87	3.13	+.26
Astr410	2.93	2.83	-.10	Astr320	2.90	2.93	+.03
Avg:	2.86	2.77	-.09	Avg:	2.94	3.00	+.06

Yet another way of looking at these data involve taking the separate ratings of outcome contributions and looking at the relationship of these ratings to particular curricular characteristics of the courses (viz., what students were actually required to do to earn their grades). For the most part the results of this analysis were disappointing; they revealed no simple, powerful pedagogy that would assure attainment of any of the outcomes. Pedagogical practices assumed to enhance learning and increase curiosity such as the provision of alternative ways to satisfy course requirements (viz., choice) or the opportunity to reaccomplish and resubmit deficient work without penalty were found to have only negligible and inconsistent effects on assessed outcome contributions. The one notable exception was the use of group work; unfortunately group work was found to have a significantly negative influence, particularly on the development of integrated knowledge and intellectual curiosity. Core courses in which a significant portion of the course depended on group work, received very low ratings for their contributions to both intellectual curiosity and integrated knowledge. The curricular characteristics found to have significant relations with any of the three educational outcomes are shown in Table 12-1.

The Influence of Curricular Characteristics on 3 Educational Outcomes (Pearson Correlation Coefficients)

Curricular Characteristics	Integrated Knowledge	Framing & Resolving	Intellectual Curiosity
% Group Work	-.50**	-.23	-.51**
% Computation	.30	-.46**	-.22
% Final	.44**	-.02	.36*
% Graded Review	.34*	-.04	.11

Table 12-1
(N = 35; *p < .05; **p < .01)

The results concerning group work run counter to a great deal of educational research literature which supports the value of collaborative learning. However, in USAFA core courses, group tasks are most likely to be employed in upper division technical courses, especially ones that serve as non-technical versions of core requirements. (In fact, the only significant relation between the proportion of group work assigned and any of the educational outcomes to be found within any of the academic divisions, was in the Engineering Division.) Perhaps under these conditions, the empowerment groups typically provided to students only serves to heighten adversarial relations between students and faculty. It is also possible that inattention to group formation and process issues may create conditions for students in which dealing with one another obscures their ability to focus on the course content. It should also be remembered that this study considered only core courses; it is possible that in majors courses, where collegial relations among students and teachers are more likely, groups may have a positive effect on the attainment of educational outcomes.

Almost by definition, emphasis on computation and arrival at singular "approved solutions" would seem to run contrary to the development of the ability to frame and resolve ill-defined problem. The student is required to assume that the problem as presented is well-defined; creative interpretations and interesting alternatives are unlikely to receive much credit. This perhaps explains why the proportion of computation involved in the course was a positive predictor of the amount of knowledge gained but a significant negative predictor of both intellectual curiosity and the ability to frame and resolve ill-defined problems. Finding a way to "contextualize" computations might be a way to avoid some of computation's disadvantages while still retaining the pedagogical advantages it offers to integrated knowledge.

Nearly all core courses require final examinations which generally range between 25% and 35% of the total grade. However, two of the courses mentioned above which involved extensive group projects in lieu of a final examination were clear outliers. Their low ratings on knowledge and curiosity thus contributed significantly to the observed positive correlation between final exam grades and assessed course contributions to knowledge and curiosity. It is likely that these correlations are more artifactual than actual.

In summary, this assessment project rested upon a great deal of work previously accomplished by many different faculty groups. The three educational outcomes themselves appear to be relevant, measurable and entirely appropriate to the Academy's stated mission of developing and inspiring air and space leaders with vision for tomorrow. The results of this study suggest that the core courses as a group do make substantial contributions to each of the outcomes but that there is also considerable variability and thus room for improvement.

Chapter 13

Conclusions and Implications for Unit Self Assessment, Curricular Reform and Further Research

CHAPTER THIRTEEN -

CONCLUSIONS AND IMPLICATIONS FOR UNIT SELF ASSESSMENT, CURRICULAR REFORM AND FURTHER RESEARCH

Conclusions

The last three chapters presented a meta-analytic response to three questions: 1) what have we learned about our ability to assess our core courses' contributions to three educational outcomes; 2) what have we learned about these core course contributions themselves; and 3) what has been learned about the three educational outcomes? This final chapter begins with a brief recap of the responses to each of these questions contained in the three previous chapters.

The reports presented in Chapters 3 through 9 showed considerable variety in approach; some reflect mere stylistic alternatives; others show more substantial distinctions. A post hoc critique yielded evaluations of the reports themselves ranging from *excellent* to *seriously flawed.* The results of two reports, in fact, required adjustment before their data was included in the data base for subsequent meta-analysis. Educational Outcomes Assessment is not free; resources are required. Assessment requires experienced and expert assessors; the nearly two years of work invested in Phase I of the Educational Outcomes Assessment Working Group appears to have provided the intellectual capital necessary for success in many of the groups. Successful assessment also requires time. Legitimate concern about imposing superfluous "busy-work" on an already over-obligated faculty led to the stipulation that no department would be required to provide more than two person hours per week to this assessment project. While this was sufficient in some cases, it was insufficient to allow for the training and education needed to prepare all participants to contribute meaningfully to their groups efforts. As a result, the level of effort in both successful and unsuccessful teams tended to be high but extremely variable across individuals. Another resource essential to success is senior level support; assessment is tough enough under ideal conditions; when the *real* expectancies of department heads are unclear, the level of task ambiguity can become overwhelming. The trust of those responding to questionnaires and critiques is also essential; self reports are notoriously susceptible to self-serving biases and rationalizations. Despite providing clear evidence of the extraordinary commitment needed to actually assess an expansive general education program such as USAFA's core curriculum, this report itself provides evidence that "it can be done". Therefore, the following conclusion seems appropriate: ***Under certain conditions, faculty teams can provide assessments of core course contributions to educational outcomes of sufficient accuracy, validity and reliability to warrant further analysis and inquiry and justify the requisite investment of institutional resources.***

The table which introduces Chapter 11 shows the combined and adjusted assessments of each core course's contribution to each of the three educational outcomes. Variation across

courses, academic divisions and educational outcomes is apparent. Such diversity is an essential prerequisite to increased understanding and eventual systemic improvement. (Without variation, there simply are no *variables* with which to work, only *constants).* The identification of such differences is no mean feat; most courses involve over a dozen instructors teaching over 50 sections of students annually. Recognizing the similarity in students, departmental faculties, and the similarity of institutional and course policies which apply equally across the curriculum; some might have predicted that differences between courses would be obscured by random variation between sections within courses. Moreover, the pattern of results reflected in Chapter 11 correspond with the *common sense* impressions of both students and many faculty members. These patterns seem to support the following generalizations about core course contributions: ***Technical courses tend to stress knowledge and convergence on approved solutions; Humanities and Social Science courses place more emphasis on thinking skills and attitudes than "facts"; non-technical versions of technical courses contribute the least of any of the core courses to any of the three educational outcomes investigated; and the "social science option" which allowed some students to select which core course to take to satisfy a single curricular requirement appears to have been very successful.***

Chapter 12 dealt with what had been learned about the three educational outcomes: integrated knowledge, the ability to frame and resolve ill-defined problems and the sustenance of intellectual curiosity. Overall, it seems fair to conclude from the available data and the teams' assessment of its significance, that the current program of core courses do contribute to all three educational outcomes. The College BASE results also support the conclusion that Air Force Academy seniors are among the most knowledgeable and broadly capable of any undergraduates in the nation. The Chapter also suggests that these three educational outcomes are somewhat distinct; especially integrated knowledge and the ability to frame and resolve ill defined problems. One somewhat surprising finding was the unexpectedly large negative influence of the use of groups on course contributions to the outcomes, particularly in upper division engineering courses. The overall pattern of results also makes clear the central importance of course contributions to students' intellectual curiosity. The following conclusions seem appropriate: ***The three educational outcomes provide a meaningful and measurable framework for educational program assessment. Although the current core curriculum contributions to all three outcomes appear to be adequate and relatively balanced, increasing emphasis on critical thinking, commitment and cognitive flexibility may require the consideration of adjustments and further reforms.***

Implications for Unit Self Assessment

The Dean's tasking to academic departments for the Spring '97 semester is to assess one core course in terms of the same outcomes assessed by the EOAWG teams. The central focus of this Unit Self Assessment (USA) is to show how assessment data are used within a department to make curricular changes to courses. As part of this tasking, departments are expected to re-assess their selected core course to validate the EOAWG results contained within this report. It is important for departments to recognize that the data upon which the individual reports and hence the meta-analysis were primarily collected during the '95-'96 academic year. After validating past core course contributions, departments should consider subsequent changes to their respective

courses and what effects, if any, these may have had on any of the three outcomes. As a result, this report can serve as an education and development tool as well as a source of data to aid in the preparation of a departmental USA.

In preparing their USA, a department should pay close attention to the assessment data contained in this report concerning their selected core course. They should also read the narrative comments concerning the assessments. Some teams' assessments are more valid than others. The importance of systematic and objective inquiry as well as logical organization and clear rationale are immediately apparent after reviewing the various reports. The appendix also contains the data considered by the teams in making their assessments. A department should include the data they feel are important, along with any other assessment data they collect on their own, in section 2.1 of their USA. Section 5.1 should describe how these data are used in their curricular review process. Comments concerning differences between the EOAWG's conclusions and their own assessment also should be made in section 5.1. Implications for curricular improvements based on these assessments should be highlighted and incorporated in departmental plans.

In the future, a department can also expect their USA to include assessments of additional core courses, their majors program and service courses. This Phase II Report has many examples (both good and bad) that can be used to help guide these efforts; there's no reason for us as an institution to repeat our mistakes. The main implication is that educational assessment is not a task that can be conducted quickly or in a ballistic manner (viz., without continual guidance and feedback). It took two years for the EOAWG's most effective assessment team members to develop the knowledge and skills they needed to conduct a valid assessment. Departments need to realize they need to invest adequate resources to obtain useful results. The development of effective educational assessment programs is now a critical aspect of the North Central Association of Colleges and Universities criteria for accreditation. Assessment is more than a passing fad. Each department needs to establish an ongoing and systematic assessment process that is deeply incorporated into their curricular processes.

Even beyond assessment of course contributions to educational outcomes, Departmental Unit Self Assessments can use this report as a general model for academic process assessment. Basically, processes are undertaken because they are believed to contribute in some way to the outcomes an organization values. Clearly stating these expected contributions establishes the assessment criteria. Systematically gathering data which might disconfirm this relationship is the essence of good assessment as well as good science. Processes such as human resource (i.e., faculty and staff) acquisition and development could readily be adapted to the general model presented in this report. The framework developed within educational assessment generalizes to the assessment of any process. Assessment is an iterative process; gradual progress depends on persistent efforts to improve effectiveness by increasing understanding and trust simultaneously.

Implications for Curricular Reform

Philosophy: Assessment is primarily the means through which an educational institution comes to better understand itself and its effect on its students, their knowledge, their skills and their attitudes. Assessment includes both evaluation of contributions (*how much* or *how well*) and

diagnosis (*by what mechanism* or *through what means*). There is danger in adopting the simplistic approach of using assessment data to reward or punish departments (or institutions) rather than to understand educational processes. Such reductionistic attempts at instant system improvement would undoubtedly destroy program coherence and also quickly undermine the trust upon which effective assessment depends. Curriculum decisions are a complex process; many factors must be considered. A course's assessed contributions to educational outcomes are not irrelevant to such decisions. However, many other factors must be considered and will often take precedence over assessed contributions.

Colonel Randy Stiles (DFE) suggests five questions critical to curricular review. The first concerns the course's purpose and the second its place or function within the overall curricular system. These first two questions are largely independent of a course's actual or assessed contributions to an institution's desired educational outcomes. Their focus is more on potential and conceptual coherence than actual educational performance. These questions are likely to be strongly influenced by factors outside the institution (i.e., the existing qualifications of students, the type and cost of available technology and the anticipated requirements of graduates). They also reflect the necessity of considering the whole program (all the outcomes and all the contributions of all the courses all at once) not just individual courses in isolation.

Colonel Stiles' other three curricular review questions depend upon assessment data: 3) how does this course affect student motivation and intellectual curiosity?; 4) to what degree does this course develop cadets' knowledge and skills?; and 5) how much is enough? A curriculum cannot be constructed entirely of promises and good intentions; great potential must eventually result in demonstrated performance. Curricular adjustments must be made on observed effects and demonstrated contributions, not on pretense, "logical" persuasion or political maneuver. Although assessment data is merely a snapshot, it can have a powerful impact on curricular reforms. If used as a source of potentially disconfirming evidence, assessment data can be used to distinguish between promises and performance; it can provide viable frames for distinguishing between curricular alternatives and even suggest potential soft spots which may require additional inquiry and/or support.

Even if one accepts the propriety of considering assessment data as a legitimate part of the curricular reform process, there is reason to be cautious about using the data contained in this report. The Phase II project concerned only three of seven educational outcomes (it is possible that some courses contribute a great deal to the outcomes not assessed in this report). This project also relied upon data from a variety of sources and instruments that are still undergoing development, validation and refinement. Additionally the decisional processes used by the teams were not fully completed (assessment teams did not have the opportunity to meet face to face and reconcile there rating differences). Under these circumstances, it is perhaps most appropriate to simply employ the results and conclusions to raise questions, concerns or highlight apparent inconsistencies for more focused and deliberate analysis rather than as final arbiters of competing curricular claims.

General Description of USAFA Core Curriculum: Before presenting recent curricular changes, it is important to provide a very brief and very general description of the core curriculum at the Air Force Academy (especially in comparison to other undergraduate curricula). Three of the

most distinctive characteristics of the Academy's Core Curriculum is that it is very large, very constrained and very technical in comparison to other undergraduate program. Of the 150+ semester hours required for most disciplinary majors, over 96 hours are devoted to core academic courses (this does not include additional required courses in physical education and military arts and sciences). In contrast, most academic majors programs at the Academy contain only 36 to 42 semester hours. At other colleges and universities, total semester hour requirements tend to be considerably smaller (130 semester hrs) and the proportion of core or general education to academic majors courses is reversed (general education requirements are typically less than one third of the total requirements).

The Air Force Academy core curriculum is also considerably more constrained than other general education programs. Most general education programs offer students alternative ways to satisfy core requirements (viz., a variety of "equivalent" courses or course sequences). In fact, some colleges and universities have recently discovered that efforts to provide flexibility and choice have resulted in general education requirements with so many alternative paths, that it is possible for two students to complete the same general education requirements without having a single course in common. This is not a problem with core course requirements at the Air Force Academy. Exceptions to the rule that all cadets take the same core courses are relatively few. A few of the most important exceptions are: the two-course foreign language requirement; technical and non-technical versions of technical core course requirements and the Social Science Option for technical majors. The Social Science Option included Macroeconomics (Econ 310), Leadership (BehSci 310) and Law for Commanders (Law 420). Cadets selecting technical majors were allowed to select one of these three courses to satisfy a single core requirement. There have also been a limited number of "honors" versions of core courses, and from time to time interdisciplinary pairings or experimental versions of core courses are offered, but these account for a very small proportion of the core course sections offered. Some curricular variety is also introduced by the fact that some cadets transfer academic credits to the Academy or validate certain basic core courses.

Another distinctive characteristic of the Academy's core curriculum is its technical orientation. Not only does the core contain a robust sequence of 24 semester hours of mathematics and physical sciences, it also contains an additional 21 semester hours of engineering courses. Air Force officers are called upon to function in a highly technical environment; it is critical that they be technologically cognizant; they must be "numerate" as well as "literate". It is also critical that as Air Force officers, graduates be aware of the social sciences; the 19.5 required semester hours of social sciences are much more than are required of students enrolled in more narrowly focused science and engineering programs.

Significant Recent Changes to the Core Curriculum:

-Elimination of Open Option requirement: In an effort to make additional room in the curriculum for more extensive flight training (required by a change in Air Force policy), there was a sensed need to reduce the number of core courses. Previously, the Dean's policy was that each major would include at least one completely open option to allow all cadets to take a course simply for the joy of learning (or whatever other reason). Some cadets used this option as intended; for others it provided the option to change majors with less of an overload. However,

in an effort to reduce the size of the core (and since this open option was included in every academic major, it was a de facto core requirement), this was the first requirement to be deleted. Assessment data concerning which particular courses were used to fulfill this requirement or what the course's relative contribution to the new educational outcomes were considered. A reasonable argument could be made that students' intellectual curiosity would be enhanced to some extent by taking a course they had freely chosen as compared to taking yet another core course from an extensive list of specific core requirements. However, deleting this requirement was attractive for another reason: deleting an open option would not have a large impact on any particular department or even a particular academic division. Perhaps in light of the potential importance of "choice" to cadets' sustenance of intellectual curiosity found in previous EOAWG studies and the current core curriculum's absence of *bona fide* curricular choices, it may be time to consider re-instituting the open option as part of the core. If nothing else, this might encourage departments to develop courses which would appeal to the interests of a broad cross section of cadets.

-Elimination of Social Science Option: The separate impacts of the dissolution of this curricular triad will be discussed below. However, it is worth noting that the concept of providing several alternative means for individual cadets to satisfy a single curricular requirement may offer advantages to both the courses and the cadets. The average assessed contributions of these three social science option courses were significantly higher than the average core courses and also slightly higher than even the elevated averages of all social science core courses. Additionally, the College BASE data suggesting increases in cadets' knowledge of social science is a prominent core curricular strength. As stated above, courses' assessed contributions should not be the only consideration in curricular reform; on the other hand, the assessed level of such contributions are relevant. The abandonment of a curricular structure with such apparently salutary effects seems ill-advised. The elimination of the Social Science Option actually involved three separate actions, each of which will be presented and discussed.

-- Merging of two 3-semester hour law courses (Law 320 and 420) into a single 4.5 semester hour course: Law 420, Law for Commanders, was the single highest rated course in terms of its contributions to all three educational outcomes assessed in this report. Law 320, the American Legal System, was also assessed as being more than one standard deviation above the core average in terms of its contributions to integrated knowledge and its overall contributions. It would seem that both courses might serve as excellent models for other courses to emulate rather than as candidates for curricular experimental surgery. On the other hand, if this combination were made to enhance or increase course coherence or accentuate integrated knowledge, such a merger might be beneficial. If this were the case, however, a more appropriate approach might be to attempt to develop such a hybrid course through a series of experimental prototypes before making more permanent curricular adjustments. Once the efficacy of the approach was demonstrated, the curriculum could be altered to accommodate the change.

-- Elimination of Macroeconomics (Econ 310): This course was rated as a solid contributor to all three educational outcomes. Alternatives to elimination might have included offering this course as an alternative to the remaining economic core course (Microeconomics) or combining the two courses as was done with the two law courses. The elimination of a social science course, making above-average contributions in all areas, is not supported by these

assessment results. The alternative of prototyping, assessing and then deciding on appropriate curricular adjustments again seems a more appropriate process for curricular adjustment.

-- **Inclusion of Leadership (BehSci 310) as core:** This course received nearly the same ratings as Economics 310 from the assessment teams. However, external emphasis and reports from recent graduates supported its inclusion in the Core For All (CFA). Taken together, these two changes (in Economics and Behavioral Sciences) appear to offset one another. However, they also constitute the elimination of one of the few actual core curricular choices some cadets previously had. These changes also represent a reduction in the previous overall breadth in the social sciences courses available in the core.

- **Creation of an "Enhanced Core":** Changes to the core curriculum replaced the Social Sciences Option with an "Enhanced Core." With this approach, each disciplinary and divisional program identifies two of four three-semester-hour courses as being required for all students enrolled in their respective majors programs. The stated functions of the Enhanced Core are (that it): provides a vehicle and process for courses to enter or depart the Core For All (CFA); provides time for assessment of those courses; comes at no cost to existing majors; and provides shared learning for each discipline beyond the CFA. Perhaps a more apt description, however, is the vernacular description, "the academic penalty box." The four courses awarded this status included two that had formerly been part of the CFA (Behavioral Science 110 and Civil Engineering 310), one that was being created in response to external Air Force interest and emphasis and one that had formerly been a requirement for most "technical" programs (Math 243). The first three of these courses will be discussed below as their changed status will result in substantial changes in their enrollment; the status change for the fourth course will not change enrollment significantly. Several differences between the Enhanced Core and the Social Science Option are worth noting. The Social Science Option actually only provided alternatives to about half the cadets (those majoring in technical programs); the Enhanced Core applies to all cadets. However, the Social Science Option allowed *cadets* to select which course would fill the requirement; the Enhanced Core (at least initially) stipulates that the choice will be made by the department overseeing the academic *program,* not the individual student (i.e., all students in any program will be required to take the same two enhanced core courses).

- **Relegation of Psychology (BehSci 110) to "Enhanced Core" status:** Because cadets are not required to declare majors until after midway through their sophomore year, placing the freshman Psychology course in the "Enhanced Core" status also required that the course be moved from the freshman to the sophomore level. This course was assessed as making contributions to cadets' ability to frame and resolve ill-defined problems and intellectual curiosity more than one standard deviation above the core average and also as the strongest overall contributor to the three outcomes during the freshman year.. As the only social science course in the freshman year, it also would seem to be a crucial contributor to the Academy's stated educational outcome of cadets developing integrated knowledge which includes all four academic divisions. Thus, this change appears to receive little logical or evidentiary support from this assessment report.

- **Relegation of Civil Engineering (CivEng 310) to "Enhanced Core" status:** Unlike Psychology, the Civil Engineering core course received very low contribution ratings; in fact, it was rated as being more than one standard deviation below the core average for its contributions

to cadets' integrated knowledge and intellectual curiosity. However, despite these difficulties, it was also rated as being slightly above the Engineering Division's average for its contributions to students' ability to frame and resolve ill-defined problems. A point to be made here is that core course contribution assessment is relatively new. The perception of punitive action based on a single assessment is problematic (and perhaps this concern contributed to the clearly aberrant assessment ratings of this course provided initially by the divisional assessment team). To the extent the Enhanced Core provides a curricular "place" for rehabilitation and renewal it would be appropriate to re-assess Civil Engineering's contributions before taking further action. It is also noteworthy that two other courses which received equally low overall ratings (Engineering 410 and Math 220) did not receive similar sanctions (i.e., being placed in the academic penalty box). (The actual assessment ratings contained in this report were not available at the time these curricular changes were made, but much of the raw data upon which they were based have been available for some time.)

- Creation of new Meteorology/Geography course as part of "Enhanced Core": This assessment report focused on the contributions of existing core courses. There is, however, at least one bit of data relevant to this course: cadets level of geographic knowledge, both at entry and graduation. The proportion of freshmen who score in the "High" category on the geography portion of the College BASE is 66%, greater than any of the other 23 subject area scores. Among seniors the proportion is 70%, tied for third highest with "political and economic structures". It should be noted that an introductory geography course has been a relatively popular option for several academic majors for several years and several Military Arts and Sciences, History and Political Science courses also address geographic knowledge and skills. In comparison to other college students, cadets appear to already be relative well-informed.

- Increase in Math 141 from 3.0 to 4.5 semester hrs: The College BASE Mathematics average score of 382 for freshmen and 406 for seniors place both groups in the 99th percentile nationally. At least within the sensitivity and range provided by the College BASE, mathematics is a relative strength of our students compared to students at other colleges and universities. If core course semester hours are a limited resource, the increased investment in an area which is already our students' greatest strength might warrant closer scrutiny. On the other hand, as stated previously, the College BASE does not include all the mathematical skills likely to be required of some Air Force officers. It may be appropriate to invest such scarce resources if they make a clear contribution to critical knowledge likely to be required of future military professionals.

Summary of Relevance of Assessment Results to Recent Curricular Changes: One approach for using assessment data for curricular improvement would be to clearly establish baselines; introduce changes under controlled conditions; and assess effects by comparing ratings of experimental groups by comparison to both control groups and previous ratings. However, real life is never as simple as laboratory formulations require and some conditions which should be controlled are almost always beyond the purview of the experimenter or assessor. Nonetheless, the results of this report may still be relevant to curricular reform in several ways. Portions of this report and the raw data it contains can be used to establish statistical baselines and also identify statistical "outliers" (which may provide an indication of the most promising areas for closer examinations). The instruments and methods used in this assessment report can also be applied to more specific assessment questions or used to compare alternative versions of a particular course

or program sequence. Of at least equal importance, is the general educational model, principles and conclusions this report contains. The review and comment on recent curricular changes above shows little evidence that core course contributions to the three educational outcomes determined any particular course's status in the core curriculum. Admittedly, many factors other than contributions to these three outcomes are likely to affect curricular decisions. However, one might expect to find more evidence that the differential level of contributions noted in this report would have served to inform the curricular changes. Although these assessment reports were accomplished after most of these curricular decisions listed above were reached, much of the data upon which these reports were made has been available for several years (e.g., course critique data, College BASE data, and curricular characteristics data). One of the key contributions of this assessment report is the collection and consolidation of these disparate sources of data.

Implications for Future Curricular Changes: Despite the fact that courses' assessed contributions appeared to have had little influence on these recent curricular decisions, this report does beg for certain curricular questions to be addressed. These include:

1. Would it be possible to decrease the number of required courses without significantly decreasing cadets' level of performance on any of the educational outcomes?

2. Would it be possible to compress or combine other courses (such as the Law 320-410 example discussed above) in such a way as to enhance educational efficacy?

3. Is there a more efficient way to achieve the desired outcomes of the current non-technical versions of technical core courses? Options might include recombining these courses with the technical versions, providing self-study validation guides for students who would prefer this method to satisfy the core requirement, grouping these courses together into a structure similar to the former Social Science Option and allowing students to choose a subset of the total.

4. Are there particular core requirements that might be better fulfilled by offering alternative but equivalent courses (viz., not strong and weak ones)? Examples might include micro-economics and macro-economics, or psychology and sociology, or civil and mechanical engineering.

5. Could broad, interdisciplinary, introductory courses in the social sciences and engineering divisions be developed and integrated into the freshman core requirements? Such courses would help establish the four-divisional integrated knowledge foundation espoused in the first educational outcome. It would also enable cadets to make more informed choices concerning the core options and alternatives mentioned above as well as decisions concerning eventual academic majors. The current experimental, problem-based engineering course would seem to be a very appropriate prototype for such integrative foundational courses.

6. Would it be useful to explore the development of interdisciplinary core courses which focus directly on the development of certain skills specified by the educational outcomes? For example a course in critical thinking and problem solving might focus on enhancing students' ability to frame and resolve ill-defined problems across disciplines.

7. Would it be possible to allow cadets to satisfy the core requirements to develop advanced skills in certain areas through the completion of majors' courses rather than core courses? One example might be statistics, as taught in Math 220/300; many of the social science majors include their own discipline-specific quantitative methods courses which cover much of the same material in even greater detail. Another example might be the advanced speaking and writing skills taught in English 311; rhetoric, argument and report writing skills might all lend themselves to disciplinary exploration and are already included as a significant part of many existing majors programs.

8. Is it time to reconsider the use of student groups in core courses? There is a great deal of evidence in the literature that groups can be a potent pedagogy for enhancing student learning, enjoyment and retention of academic material. However, this report contains a great deal of evidence that, as presently employed in core courses, group work may be detrimental to students' attainment of all three educational outcomes. Although not presented in this report, substantial evidence is also available that student groups have made a myriad of positive contributions to student education within majors programs. A hiatus on group work in core courses would allow what we've learned in majors courses to be applied to the use of groups in core courses.

Implications for Further Research

This report builds upon the previous research conducted by the Educational Outcomes Assessment Working Group as reported in two previous documents: The EOAWG Phase I: Inital Report and EOAWG Phase I: Final Report. Together, these three documents provide a semi-structured literature and record of systematic inquiry into basic educational research at the Air Force Academy. There have also been a variety of other research activities which have provided valuable elaboration and insight concerning many different educational practices, processes and programs. It seems appropriate to close this assessment report by specifically identifying several of the many opportunities for further educational research suggested by these reports. A sample of some of these questions follow.

The tendency of students' course ratings to intensify with time is relatively well established by many different educational researchers, including Joseph Petit, William Cashin and Wilbert McKeachie. In this context, Engineering 410, the engineering capstone course, appears to provide a striking counter-example. This course seems to consistently receive relatively negative end-of-course critiques; however, it is just as consistently identified by junior officers as being among the most relevant to a variety of activities they are required to perform as junior officers in the Air Force. A plethora of anecdotal accounts consistently also endorse the valuable contributions of this course to student's ability to frame and resolve ill-defined problems and to learn to work effectively together. Other courses which receive similarly low end-of-course ratings but do not appear to undergo such a post hoc evaluation reversals are Civil Engineering 310 and Mathematics 220. Understanding why and how the reversal occurs in Engineering 410, but apparently not in these other courses, might significantly enhance our understanding of basic educational processes and their contributions to educational outcomes.

The failure of this report to identify curricular characteristics which were significant predictors of the level of contributions to the three educational outcomes poses a significant challenge. One approach might be to attempt to identify variation within core courses which significantly predict differential contributions. Another approach might be to attempt developing alternative dimensions of difference between core courses which might be reasonably expected to predict differential contributions. The use of groups, degree of choice in satisfying course requirements and opportunity to reaccomplish deficient work without penalty are all curricular characteristics supported by contemporary research in higher education which did not receive significant support from this assessment. One possible explanation for this is that all three practices are at such a low level in core courses that the restricted range did not provide significant statistical leverage. Obviously this problem would not apply to the use of groups which did reach the level of statistical significance but in a direction opposite to the one expected. A somewhat related question involves identification of specific data that were most closely related to teams' assessment ratings. For example, some faculty members have suggested that horizontal teams may have relied more heavily (or even exclusively) on student critique data. The question of which particular items were most predictive of eventual outcomes might help to streamline future assessment efforts.

The discovery of the apparently negative correlation between group work and contributions to all three educational outcomes was somewhat surprising (but not unprecedented). Ted Marchese of the American Association of Higher Education, along with many other noted authorities, lists collaborative and cooperative learning as being among the most powerful and promising pedagogies in higher education. However, a review of much of the literature supportive of the value of educational groups raises considerable suspicion of potential experimenter bias; many demonstrations by ardent group learning advocates (such as the Johnsons or Alfie Kohn) appear to be confounded by the self-fulfilling prophecy. Other research at the Academy seems to corroborate the concern for the potentially negative influence of groups. Since the negative influence of group work appears to be most concentrated in upper division engineering courses, these courses might be the best place to start to systematically explore this paradoxical phenomenon. It might also be possible to identify positive examples of group work's potential positive contributions to educational outcomes by considering the ways groups are used in courses within various academic majors.

This assessment report suggests that students' intellectual curiosity is a measurable variable and consequential educational outcome. The distribution of assessed contribution ratings for this outcome was very similar to those for the other two outcomes and the meta-analysis showed significant correlations with both the other outcomes. In fact, the greatest degree of variability in this outcome appears to be in faculty members' understanding and appreciation of the importance of their students' intellectual curiosity rather than the actual level of assessed course contributions to this outcome. Opportunity for significant improvements in the core's educational effectiveness may be contingent upon developing ways to better educate the entire faculty concerning the development processes and the essential role of curiosity and intrinsic motivation. The 1996 Fall semester's faculty-wide colloquium on the subject of intellectual curiosity, its measurement and its contribution to learning, may have been a big step in the right direction. Certainly the attitudinal data collected as a part of this effort might serve as a base-line for charting the effects of future faculty development initiatives toward this end.

In addition to the fall colloquium, a research project that attempts to measure intellectual curiosity should also begin. A preliminary effort was started during the fall 1996 semester using a questionnaire developed by the Center for Educational Excellence. The questionnaire asked students to rate their agreement with several questions that attempted to capture student attitudes toward their own learning. Preliminary data indicate the Academy has a negative influence on cadet attitudes toward learning that grows through the four-year experience. These initial findings were also supported by the results of the Fall 1996 semester IDEA course critique data that indicate the Academy is far below the national average for student motivation to take courses. Research efforts to examine the validity of measuring intellectual curiosity are continuing with an attempt to integrate the findings into curricular reform discussions.

Another one of the educational outcomes which invites further investigation is our students' ability to frame and resolve ill-defined problems. The EOAWG Phase I: Final Report indicates that there is considerable room for improvement in this area with nearly 40% of our Spring semester seniors scoring below the fully satisfactory level. The lack of clearly differential abilities based on academic major (and GPA) suggests that the causes apparently undermining our cadets' ability in this critical area are likely to be found in our core rather than our majors programs. The results of the College BASE contained in this report suggest that cadets enhanced technical abilities to solve equations are to some extent offset by an apparently decreased ability to comprehend contextual aspects and broader implications of the problems. This is consistent with the "plug-and-chug" mentality cadets sometimes complain about as being imposed upon them some core courses. Further investigation of cadets' problem solving ability and how it is influenced by particular core courses is needed.

In addition, another project aimed at validating previous data on the development of cadets' problem solving ability began this semester. In this project, an attempt will be made to use the Reflective Judgment Model framework (King & Kitchner) to assess the ability of cadets to frame and resolve ill-defined problems across the four years of Academy experience. Three instruments will be used in an attempt to cross-validate the instruments and measure cadet problem solving skills. If these data confirm previous findings that are less than satisfactory, then a critical mass of data will exist that indicate the need for curricular reform.

This report tacitly assumed that the effect of any particular core course on any particular educational outcome was the same for all cadets. Informal anecdotal evidence uncovered through these assessment activities, as well as other research and climate surveys, suggests that there may be considerable variability of the impact on various cadets. For example, women seem to have considerably more adverse experiences with group work than their male counterparts. Women also are likely to perceive somewhat higher levels of gender-based faculty discrimination in the classroom than male cadets. Similarly cadet climate surveys indicate that racial and ethnic minorities are somewhat more susceptible to the perception of negative expectations by faculty members. Better understanding the differential contributions to and development of these educational outcomes among women and minority students remains a significant goal for educational research at the Academy (as well as throughout higher education).

Obviously there are many other areas for further inquiry as well. Hopefully, the tools, techniques and general tenor of this report can serve as catalysts for other investigations of the

educational process and its products (viz., attainment of all seven of the Academy's educational outcomes).

Closing Comment

This has been an extensive and intensive assessment effort. These pages document a sincere and authentic struggle to understand the influence (both individually and in combination) of core courses from across the curriculum on three of our seven educational outcomes. There has been disagreement, and misunderstanding; miscommunication and confusion throughout the process; assessment is often messy as well as fuzzy. There has been a conscientious effort to report such disagreements rather than to gloss over them. In some cases the findings may appear contradictory or counter-intuitive. In all honesty, this report has yielded many more questions than answers. There are some who will be quick to point out that the patterns and relationships identified or suggested in both the individual and meta-analytic reports are less than convincing. Others will argue that this report has provided little that was not already known, at least by some. But this report is a start: a start to a potentially substantive and enriching conversation about experimentation and improvement within our core curriculum. John Stuart Mill's masterpiece, *On Liberty*, suggests that silencing any opinion is "a peculiar evil": if the opinion were correct, its suppression denies us its direct enlightenment; if the opinion were incorrect, its suppression deprives us of the opportunity to gain insight from its collision with truth. The *sine qua non* of assessment is the same as that of science: a true effort to disprove one's own underlying (and often most cherished) assumptions. It is in that context, the ideas, opinions and conjecture contained in this report are offered.